

APPARATUS AND METHOD FOR INFORMATION PROCESSING, AND RECORDING MEDIUM AND PROGRAM

Publication number: JP2003296365 (A)

Publication date: 2003-10-17

Inventor(s): YAMAMOTO NORIYUKI; SAITO MARI +

Applicant(s): SONY CORP +

Classification:

- international: G06F13/00; G06F17/30; G06F7/00; G06F13/00; G06F17/30; G06F7/00; (IPC1-7): G06F13/00; G06F17/30

- European: G06F17/30; G06F13/00; G06F17/30

Application number: JP20020095413 20020329

Priority number(s): JP20020095413 20020329

Also published as:

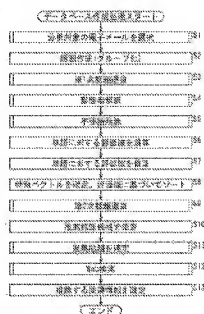
JP4082059 (B2)

US2003220922 (A1)

Abstract of JP 2003296365 (A)

PROBLEM TO BE SOLVED: To present related information corresponding to interests of a user. ; SOLUTION: An electronic mail message to be analyzed which is defined as an origin of database creation is selected in a step S1. This electronic mail message to be analyzed is sorted into the respective topics in a step S2. The topics are primarily picked up in a step S3. Linked texts corresponding to the respective picked topics are decomposed into words by morpheme analysis in a step S4. Unnecessary words are removed in a step S5. Evaluation values are imparted to the words in a step S6 and the evaluation values are corrected in a step S7. The topics are secondarily picked up in a step S8. candidates for recommended topics are established in a step S10 and recommended topics are established in a step S11. Places of the related information corresponding to the recommended topics are retrieved on the Web in a step S12. ; For example, the related information is applied to a program for displaying a desktop mascot on a screen of a personal computer. ; COPYRIGHT: (C) 2004, JPO

図3



Data supplied from the **espacenet** database — Worldwide

(51)Int.Cl. ⁷	識別記号	P 1	チーフ・ド (参考)
G 0 6 F 17/30	3 5 0	G 0 6 F 17/30	3 5 0 C 5 B 0 7 5
	1 8 0		1 8 0 Z
	2 1 0		2 1 0 D
	3 1 0		3 1 0 A
	3 4 0		3 4 0 B
審査請求 未請求 請求項の数29 O L (全 30 頁) 最終頁に続く			
(21)出願番号	特願2002-95413(P2002-95413)	(71)出願人	000002185 ソニー株式会社 東京都品川区北品川 6 丁目 7 番35号
(22)出願日	平成14年3月29日 (2002. 3. 29)	(72)発明者	山本 則行 東京都品川区北品川 6 丁目 7 番35号 ソニ ー株式会社内
		(72)発明者	斎藤 真里 東京都品川区北品川 6 丁目 7 番35号 ソニ ー株式会社内
		(74)代理人	100082131 弁理士 稲本 義雄 Fターム(参考) 5B075 NI20 PP03 PP13 PR08

(54)【発明の名称】 情報処理装置および方法、記録媒体、並びにプログラム

(57)【要約】

【課題】 ユーザの興味に対応する関連情報を提示する。

【解決手段】 ステップS1で、データベース作成の素とする分析対象電子メールが選択される。ステップS2で、分析対象電子メールが各話題に分類される。ステップS3で、話題が1次選択される。ステップS4で、選択された各話題に対応する連結本文が形態素解析により単語に分解される。ステップS5で、不要語が除去される。ステップS6で、単語に評価値が付与され、ステップS7で評価値が修正される。ステップS9で、話題が2次選択され、ステップS10で推薦話題候補が、ステップS11で推薦話題が確定される。ステップS12で、推薦話題に対応する関連情報がWeb検索される。例えば、パーソナルコンピュータの画面上にデスクトップマスコットを表示させるプログラムに適用することができる。

図3



【特許請求の範囲】

【請求項1】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成するデータベース生成手段と、

所定の文書情報から特徴語を探索する検索手段と、前記データベース生成手段によって生成された前記関連情報のうち、前記検索手段によって探索された前記特徴語に関連するものを提示する提示手段とを備える情報処理装置において、

前記データベース生成手段は、

全ての前記既存の文書情報のうち、前記グループに分類する処理の対象とする前記既存の文書情報を選択する選択手段と、

前記選択手段によって選択された前記既存の文書情報を前記グループに分類する分類手段と、少なくとも1以上の前記既存の文書情報からなる前記グループを選択する選択手段と、

前記グループに対応する関連情報を取得する取得手段と、

前記取得手段によって取得された前記関連情報を、前記グループに対応付けて蓄積する蓄積手段とを含むことを特徴とする情報処理装置。

【請求項2】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成する情報処理装置において、

全ての前記既存の文書情報のうち、前記グループに分類する処理の対象とする前記既存の文書情報を選択する選択手段を含むことを特徴とする情報処理装置。

【請求項3】 前記選択手段は、全ての前記既存の文書情報のうち、所定の期間における通信頻度、通信日時、および通信回数などの少なくとも1つに基づいて決定する通信相手条件を満たす相手との間で通信した前記既存の文書情報を選択することを特徴とする請求項2に記載の情報処理装置。

【請求項4】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成する情報処理装置の情報処理方法において、

全ての前記既存の文書情報のうち、前記グループに分類する処理の対象とする前記既存の文書情報を選択する選択ステップを含むことを特徴とする情報処理方法。

【請求項5】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成するプログラムが記録されている記録媒体。

【請求項6】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成するコンピュータに、

全ての前記既存の文書情報のうち、前記グループに分類する処理の対象とする前記既存の文書情報を選択する選択ステップを実行させるプログラム。

【請求項7】 既存の文書情報のグループに対応する関連情報からなるデータベースを生成する情報処理装置において、

前記既存の文書情報を前記グループに分類する分類手段を含むことを特徴とする情報処理装置。

【請求項8】 前記選択手段は、構成数条件を満たさない前記既存の文書情報からなる前記グループを除去することを特徴とする請求項7に記載の情報処理装置。

【請求項9】 前記選択手段は、前記グループの数に対応して前記構成数条件を変更することを特徴とする請求項7に記載の情報処理装置。

【請求項10】 既存の文書情報のグループに対応する関連情報からなるデータベースを生成する情報処理装置の情報処理方法において、

前記既存の文書情報を前記グループに分類する分類ステップを含むことを特徴とする情報処理方法。

【請求項11】 既存の文書情報のグループに対応する関連情報からなるデータベースを生成するプログラムであって、

前記既存の文書情報を前記グループに分類する分類ステップを含むことを特徴とするコンピュータが読み取り可能なプログラムが記録されている記録媒体。

【請求項12】 既存の文書情報のグループに対応する関連情報からなるデータベースを生成するコンピュータに、前記既存の文書情報を前記グループに分類する分類ステップを実行させるプログラム。

【請求項13】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成する情報処理装置において、

少なくとも1以上の前記既存の文書情報からなる前記グループに対応する関連情報を取得する取得手段を含むことを特徴とする情報処理装置。

【請求項14】 前記取得手段は、

同一の前記グループに分類されている全ての前記既存の文書を連結して連結文書を生成する連結手段と、

形態素解析によって前記連結文書を単語に分解する形態素解析手段と、

形態素解析手段によって分解された単語に所定の条件に従って加重した評価値を付与する評価値付与手段と、

前記グループに前記評価値が付けられた単語を要素とする単語ベクトルを設定する単語ベクトル設定手段と、前記グループに対応する前記単語ベクトルの要素である前記単語を検索語とし、ネットワーク上の検索エンジンを用いて前記関連情報を取得する検索手段とを含むことを特徴とする請求項13に記載の情報処理装置。

【請求項15】 前記連結手段は、同一の前記グループ

に分類されている全ての前記既存の文書を、送信した前記既存の文書と受信した前記既存の文書との間に所定の文字列を挿入して連結し、前記連結文書を作成することを特徴とする請求項14に記載の情報処理装置。

【請求項16】 前記評価値付与手段は、送信した前記既存の文書に属していた単語に対し、受信した前記既存の文書に属していた単語よりも加重して評価値を付与することを特徴とする請求項14に記載の情報処理装置。

【請求項17】 前記評価値付与手段は、前記単語に対し、前記単語が属している前記既存の文書の数および長さの少なくとも一方に対応して加重した評価値を付与することを特徴とする請求項14に記載の情報処理装置。

【請求項18】 前記単語ベクトル設定手段は、単語ベクトルから不要語を削除することを特徴とする請求項14に記載の情報処理装置。

【請求項19】 少なくとも1以上前記既存の文書情報からなる前記グループを選択する選択手段をさらに含む。

前記選択手段は、対応する単語ベクトルの要素の数が所定の数以下の前記グループを除去することを特徴とする請求項14に記載の情報処理装置。

【請求項20】 前記単語ベクトル設定手段は、前記選択手段により、対応する単語ベクトルの要素の数が所定の数以下の前記グループが除去された結果、選択された前記グループに対応する単語ベクトルから不要語を削除することを特徴とする請求項19に記載の情報処理装置。

【請求項21】 前記選択手段は、前記単語ベクトル設定手段によって前記不要語が除去されたことにより、対応する単語ベクトルの要素の数が所定の数以下となった前記グループも除去することを特徴とする請求項19に記載の情報処理装置。

【請求項22】 前記評価値付与手段は、前記単語ベクトル設定手段によって前記単語ベクトルから前記不要語が削除され、かつ、前記選択手段によって対応する単語ベクトルの要素の数が所定の数以下の前記グループが除去された後、前記単語に対して、前記所定の条件に従って加重した前記評価値を付与することを特徴とする請求項19に記載の情報処理装置。

【請求項23】 前記選択手段は、対応する単語ベクトルの要素である単語に付与されている評価値の最大値が所定の値以上であって、かつ、分類されている前記既存の文書の最新の通信日時が所定の期間内である前記グループも除去することを特徴とする請求項19に記載の情報処理装置。

【請求項24】 所定の文書情報から特徴語を検索する検索手段をさらに含む。前記検索手段は、前記グループに対応する前記単語ベクトルのうち、付与されている前記評価値が上位の複数の単語を連結して検索語とすることを特徴とする請求項14に記載の情報処理装置。

【請求項25】 前記検索手段は、前記検索エンジンから取得した検索結果のうち、所定の文字列を含むものを前記関連情報から除外することを特徴とする請求項24に記載の情報処理装置。

【請求項26】 前記検索手段は、予め設定されている単語を検索素とすることを特徴とする請求項24に記載の情報処理装置。

【請求項27】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成する情報処理装置の情報処理方法において、少なくとも1以上の前記既存の文書情報からなる前記グループに対応する関連情報を取得する取得ステップを含むことを特徴とする情報処理方法。

【請求項28】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成するプログラムであって、少なくとも1以上の前記既存の文書情報からなる前記グループに対応する関連情報を取得する取得ステップを含むことを特徴とするコンピュータが読み取り可能なプログラムが記録されている記録媒体。

【請求項29】 既存の文書情報をグループに分類し、前記グループに対応する関連情報からなるデータベースを生成するコンピュータに、少なくとも1以上の前記既存の文書情報からなる前記グループに対応する関連情報を取得する取得ステップを実行させるプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報処理装置および方法、記録媒体、並びにプログラムに関し、特に、電子メールなどの文書の中から、ユーザの興味があると思われる単語および関連情報を取得してデータベースに蓄積し、その関連情報を効果的に表示させるようにした情報処理装置および方法、記録媒体、並びにプログラムに関する。

【0002】

【従来の技術】従来、パーソナルコンピュータのデスクトップ（表示画面）に、いわゆるデスクトップマスコットと呼ばれるキャラクターを表示させるアプリケーションプログラムが存在する。

【0003】デスクトップマスコットは、例えば、電子メールの着信等をユーザに通知する機能やデスクトップ上を移動する機能を有している。

【0004】ところで、例えば、ユーザが電子メールとして送信する文書等を入力している時や、受信した文書を開覧している時などにおいて、送受信の対象としている文書に関連する情報（以下、関連情報と記載する）をユーザに提示することができれば、ユーザによって利便性が向上する。さらに、当該提示をデスクトップマスコットが実行するようになれば、デスクトップマスコット

に対して一層要着を感じるようになると考えられる。

【0005】従来、電子メールなどの文書を用いて自動的にデータベースを構築し、送受信した電子メールの文書に関連する関連情報をユーザに提示する方法が、例えば特開2001-312515号公報(以下、先願と記述する)に開示されている。

【0006】

【発明が解決しようとする課題】しかしながら、先願の発明においては、電子メールの利用状況の個人差、すなわち、電子メールの使用歴の長短、送受信頻度の高低、フォルダ分類の有無、通信相手の多少などが考慮されることなく、全ての電子メールが分析されてデータベース化が行われていたため、その分析処理にコンピュータのリソース(処理時間、メモリ等)を浪費してしまうことが多分に生じていた。また、分析結果が適切ではないことが多く、ユーザに対して適切な情報を提示することができない課題があった。

【0007】すなわち、先願においては、ユーザが興味をもっている事柄に対応する単語を電子メールの文章の中から抽出し、抽出した単語に対応する情報をユーザに提示するようにしていた。ユーザの興味に対応する単語を電子メールの文章の中から抽出する方法は、具体的には、ユーザの興味に文章中に使用する単語の出現頻度に影響するという仮定に基づき、全ての電子メールまたは一定期間に通信した電子メールを対象として、1通毎に形態素解析を実行して単語を抽出し、抽出した各単語の出現頻度を計算し、1通毎あるいは一定期間に通信した複数の電子メール毎に、出現頻度が高い単語を、ユーザの興味に対応する単語として抽出するようにしていた。

【0008】しかしながらこのような従来方法では、電子メールの利用状況の個人差、および電子メールが有している特徴(例えば、送受信者、通信日時などを特定できること)を全く利用していないので、例えば、受信しても返信することがないメーリングリストからの電子メール、宣伝用のいわゆるスパム電子メールなども分析の対象としてしまい、ユーザの興味には関係がない単語が抽出されてしまうことがあった。

【0009】また、従来方法では、送受信した電子メールを分析の対象としていたため、電子メールが送受信されない状況においては、ユーザの興味に対応する単語が新たに抽出されない状況においても、新たな関連情報をユーザに提示することができない課題があった。

【0010】なお、従来、ユーザの興味に対応する単語が新たに抽出されない状況においても、何らかの情報をユーザに提示するために、一般的な情報を表示するWebページのURLとタイトルを予め登録しておく方法が存在する。しかしながら、この方法では、ユーザの興味に対応する単語が新たに抽出されない状況において、毎回同一のWebページが提示されることになるので、ユーザにとって意外性がないだけでなく、当該WebページのURLが

変更された場合、それに対応することができない課題があった。

【0011】本発明はこのような状況に鑑みてなされたものであり、電子メールの特徴に基づき、分析する文章を限定することにより、速やかにユーザの興味に対応する単語を抽出できるようにするとともに、電子メールの送受信が行われない状況においても、ユーザに適切な情報を提示できるようにすることを目的とする。

【0012】

【課題を解決するための手段】本発明の第1の情報処理装置は、データベース生成手段は、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報を選択する選択手段と、選択手段によって選択された既存の文書情報をグループに分類する分類手段と、少なくとも1以上の既存の文書情報からなるグループを選択する選択手段と、グループに対応する関連情報を取得する取得手段と、取得手段によって取得された関連情報を、グループに対応付けて蓄積する蓄積手段とを含むことを特徴とする。

【0013】本発明の第2の情報処理装置は、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報を選択する選択手段を含むことを特徴とする。

【0014】前記選択手段は、全ての既存の文書情報のうち、所定の期間における通信頻度、通信日時、および通信経数のうちの少なくとも1つに基づいて決定する通信相手条件を満たす相手との間で通信した既存の文書情報を選択するようにすることができる。

【0015】本発明の第2の情報処理方法は、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報を選択する選択ステップを含むことを特徴とする。

【0016】本発明の第2の記録媒体のプログラムは、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報を選択する選択ステップを含むことを特徴とする。

【0017】本発明の第2のプログラムは、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報を選択する選択ステップをコンピュータに実行させることを特徴とする。

【0018】本発明の第3の情報処理装置は、既存の文書情報をグループに分類する分類手段を含むことを特徴とする。

【0019】前記選択手段は、構成数条件を満たさない既存の文書情報からなるグループを除去するようにすることができる。

【0020】前記選択手段は、グループの数に対応して構成数条件を変更するようにすることができる。

【0021】本発明の第3の情報処理方法は、既存の文書情報をグループに分類する分類ステップを含むことを

特徴とする。

【0022】本発明の第3の記録媒体のプログラムは、既存の文書情報をグループに分類する分類ステップを含むことを特徴とする。

【0023】本発明の第3のプログラムは、既存の文書情報をグループに分類する分類ステップをコンピュータに実行させることを特徴とする。

【0024】本発明の第4の情報処理装置は、少なくとも1以上の既存の文書情報からなるグループに対応する関連情報を取得する取得手段を含むことを特徴とする。

【0025】前記取得手段は、同一のグループに分類されている全ての既存の文書を連結して連結文書を生成する連結手段と、形態素解析によって連結文書を単語に分解する形態素解析手段と、形態素解析手段によって分解された単語に所定の条件に従って加重した評価値を付与する評価値付与手段と、グループに評価値が付与された単語を要素とする単語ベクトルを設定する単語ベクトル設定手段と、グループに対応する単語ベクトルの要素である単語を検索語とし、ネットワーク上の検索エンジンを用いて関連情報を取得する検索手段とを含むようにすることができる。

【0026】前記連結手段は、同一のグループに分類されている全ての既存の文書を、送信した既存の文書と受信した既存の文書との間に所定の文字列を挿入して連結し、連結文書を生成するようにすることができる。

【0027】前記評価値付与手段は、送信した既存の文書に属していた単語に対し、受信した既存の文書に属していた単語よりも加重して評価値を付与するようにすることができる。

【0028】前記評価値付与手段は、単語に対し、単語が属している既存の文書の数および長さの少なくとも一方に対応して加重した評価値を付与するようにすることができる。

【0029】前記単語ベクトル設定手段は、単語ベクトルから不要語を削除するようにすることができる。

【0030】本発明の第4の情報処理装置は、少なくとも1以上の既存の文書情報からなるグループを選択する選択手段をさらに含むことができ、前記選択手段は、対応する単語ベクトルの要素の数が所定の数以下のグループを除去するようにすることができる。

【0031】前記単語ベクトル設定手段は、選択手段により、対応する単語ベクトルの要素の数が所定の数以下のグループが除去された結果、選択されたグループに対応する単語ベクトルから不要語を削除するようにすることができる。

【0032】前記選択手段は、単語ベクトル設定手段によって不要語が除去されたことにより、対応する単語ベクトルの要素の数が所定の数以下となったグループも除去するようにすることができる。

【0033】前記評価値付与手段は、単語ベクトル設定

手段によって単語ベクトルから不要語が削除され、かつ、選択手段によって対応する単語ベクトルの要素の数が所定の数以下のグループが除去された後、単語に対して、所定の条件に従って加重した評価値を付与するようにすることができる。

【0034】前記選択手段は、対応する単語ベクトルの要素である単語に付与されている評価値の最大値が所定の値以上であって、かつ、分類されている既存の文書の最新の通信日時が所定の期間内であるグループも除去するようにすることができる。

【0035】本発明の第4の情報処理装置は、所定の文書情報から特徴語を検索する検索手段をさらに含むことができ、前記検索手段は、グループに対応する単語ベクトルのうち、付与されている評価値が上位の複数の単語を連結して検索語とするようにすることができる。

【0036】前記検索手段は、検索エンジンから取得した検索結果のうち、所定の文字列を含むものを関連情報から除外するようにすることができる。

【0037】前記検索手段は、予め設定されている単語を検索語とするようにすることができる。

【0038】本発明の第4の情報処理方法は、少なくとも1以上の既存の文書情報からなるグループに対応する関連情報を取得する取得ステップを含むことを特徴とする。

【0039】本発明の第4の記録媒体のプログラムは、少なくとも1以上の既存の文書情報からなるグループに対応する関連情報を取得する取得ステップを含むことを特徴とする。

【0040】本発明の第4のプログラムは、少なくとも1以上の既存の文書情報からなるグループに対応する関連情報を取得する取得ステップをコンピュータに実行させることを特徴とする。

【0041】本発明の第1の情報処理装置においては、全ての既存の文書情報のうち、グループに分類する処理の対象とされる既存の文書情報が選択され、選択された既存の文書情報がグループに分類される。また、少なくとも1以上の既存の文書情報からなるグループが選択され、グループに対応する関連情報が取得される。さらに、取得された関連情報が、グループに対応付けて蓄積される。

【0042】本発明の第2の情報処理装置および方法、並びにプログラムにおいては、全ての既存の文書情報のうち、グループに分類する処理の対象とする既存の文書情報が選択される。

【0043】本発明の第3の情報処理装置および方法、並びにプログラムにおいては、既存の文書情報がグループに分類される。

【0044】本発明の第4の情報処理装置および方法、並びにプログラムにおいては、少なくとも1以上の既存の文書情報からなるグループに対応する関連情報が取得

される。

【0045】

【発明の実施の形態】以下、本発明の実施の形態について、図面を参照して説明する。図1は、本発明を適用したデスクトップマスコット(以下、エージェントと記述する)をデスクトップ上に表示するためのアプリケーションプログラム(以下、エージェントプログラムと記述する)1、電子メールを送受信するためのアプリケーションプログラム(以下、メーラ(mailer)と記述する)2、および、文書作成または編集するためのワードプロセッサプログラム(以下、ワープロプログラムと記述する)3との関係を説明する図である。

【0046】エージェントプログラム1乃至ワープロプログラム3は、例えば、パーソナルコンピュータ(詳細は、図2を参照して後述する)にインストールされて実行されるものである。

【0047】エージェントプログラム1は、処理の対象とする文書の関連情報(後述)を蓄積してデータベースを構築する蓄積部11、処理の対象とする文書に対応する関連情報をユーザに提示する提示部12、および、エージェント172(図21)の表示等を制御するエージェント制御部13から構成される。

【0048】なお、蓄積部11および提示部12を、例えばインターネット上の任意のサーバに設置するようにしてもよい。

【0049】蓄積部11の文書取得部21は、メーラ2によって送受信された文書やワープロプログラム3によって編集された文書などのうち、自己が未処理の文書を取得して文書属性処理部22および文書内容処理部23に供給する。

【0050】なお、以下においては、主に、メーラ2によって送受信された電子メールの文書を処理の対象とする場合の例について説明する。

【0051】文書属性処理部22は、文書取得部21から供給される文書の属性情報を抽出し、その属性情報に基づいて文書をグループ化し、文書内容処理部23および文書特徴データベース作成部24に供給する。電子メ

特徴ベクトル

$$= (\text{単語1の評価値}w_1, \text{単語2の評価値}w_2, \dots, \text{単語}n\text{の評価値}w_n) \\ \dots (1)$$

【0056】評価値の演算には、例えば文蔵(Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley, 1989)に開示されているtf-idf法を用いる。tf-idf法によれば、話題Aに対応するn次元の特徴ベクトルのうち、話題Aに含まれる単語に対応する要素に対しては、評価値として0以外の値が算出され、話題Aに含まれない単語(頻度が0である単語)に対応する要素に対しては、評価値として0が算出される。

【0057】なお、評価値は、例えば、電子メールの送

受の頻度や回数、電子メールに含まれる単語の品詞の種類(特定の地域や名称を示す固有名称など)、送受信する相手に応じて修正される。

【0052】また、一般にここで言う話題とは、電子メールに限らず、ワープロ、エディタやスケジューラなど、その他のツールやアプリケーションソフトウェアなどから作成されるようなあらゆる文書に關して、ある関係で関連付けられた一連の文書群を指す。

【0053】文書内容処理部23は、文書属性処理部22でグループ化された文書群(話題)の本文を抽出し、形態素解析を施して、単語(特徴語)に分類する。単語は、品詞(名詞、形容詞、動詞、副詞、接続詞、感動詞、助詞、および助動詞)別に分類される。ただし、広範囲に亘って分布している単語、すなわち、例えば、大多数の文書に含まれていると考えられる単語「こんにちは」、「よろしく」、あるいは「お願いします」等の名詞以外の品詞は関連情報を検索するためのキーワード(以下、検索語とも記述する)には成り得ないので、不要語であるとしてキーワードとする対象から削除される。

【0054】また、文書内容処理部23は、不要語が削除された各単語の出現頻度および複数の文書に亘る分布状況を求め、グループ化された文書群(話題)毎に、各単語の重み(文書の主旨に関連する程度を示す値、以下、評価値と記述する)を演算する。

【0055】さらに、文書内容処理部23は、各話題に対し、単語の評価値を要素とする特徴ベクトルを決定する。例えば、各話題に含まれる単語(特徴語)の総数がn個である場合、各話題の特徴ベクトルは、n次元空間のベクトルとして次式(1)のように表現される。

受信の頻度や回数、電子メールに含まれる単語の品詞の種類(特定の地域や名称を示す固有名称など)、送受信する相手に応じて修正される。

【0058】また、本実施の形態においては、話題毎に特徴ベクトルを算出するものとして説明するが、これに限らず、文書毎、または、その他の単位毎(例えば、所定期間(1週間)に蓄積された文書群毎)に特徴ベクトルを算出するようにすることも勿論可能である。

【0059】文書特徴データベース作成部24は、文書属性処理部22によってグループ化された文書群(話

題)毎の各文書の属性情報と、文書内容処理部23で算出された話題毎の特徴ベクトル(すなわち、話題内に含まれる単語の評価値)を時系列順にデータベース化して、ハードディスクドライブなどよりなる記憶部49(図2)に記録する。また、文書特徴データベース作成部24は、単語の評価値などを参照することにより、所定の条件を満たす単語を選択し、関連情報を検索するための検索用キーワード(検索語)として記録する。さらに、文書特徴データベース作成部24は、検索語を関連情報検索部25に供給し、それに対応して関連情報検索部25から供給される関連情報を、検索語に対応付けて記録する。

【0060】関連情報検索部25は、文書特徴データベース作成部24から供給される検索語に対する関連情報を検索し、検索結果のインデックスを文書特徴データベース作成部24に供給する。検索語に対する関連情報を検索する方法としては、例えばインターネット上の検索エンジンを用いる方法がある。検索エンジンを用いる方法を適用した場合、検索結果として得られるWebページのURL(Uniform Resource Locator)とWebページのタイトルが、関連情報として文書特徴データベース作成部24に供給される。

【0061】提示部12のイベント管理部31は、メーラ2がアクティブとされること、メーラ2が電子メールの送受信を完了したこと、および、入力中の文書のテキストデータ量が所定の閾値を超えたことを検知して、データベース問い合わせ部32に通知する。以下、メーラ2が電子メールの送受信を完了したこと、または、入力中の文書のテキストデータ量が所定の閾値を超えたことを、イベント発生と記述する。

【0062】また、イベント管理部31は、内蔵するタイマ31Aを参照することによって時間の経過を監視し、直前、所定のタイミングから所定の時間が経過した場合、その旨をデータベース問い合わせ部32に通知する。

【0063】データベース問い合わせ部32は、イベント管理部31からのイベント発生時の通知に対応して、イベント発生に対応する文書(例えば、受信した電子メール)を取得し、文書内容処理部23の処理と同様に、その文書に形態素解析を施して単語を抽出し、不要語を除外して各単語の評価値を演算する。これにより、イベント発生に対応する文書の特徴ベクトルが算出される。

【0064】また、データベース問い合わせ部32は、文書特徴データベース作成部24によって作成されたデータベースを検索し、算出したイベント発生に対応する文書の特徴ベクトルと、データベースに記録されている話題毎の特徴ベクトルとの内積を、両者の類似度として算出する。さらに、データベース問い合わせ部32は、イベント発生に対応する文書に対する類似度が最も高い話題を判定し、その話題に含まれる単語のうち、評価値

が所定の条件(詳細については後述する)を満たすものを選択し、選択した単語(重要語)に対応する関連情報を、イベント管理部31を介し、または直接的に、関連情報提示部33に供給する。

【0065】関連情報提示部33は、イベント管理部31を介し、または直接的に、データベース問い合わせ部32から供給される関連情報を表示部48(デスクトップ)上に表示させる。すなわち、イベント管理部31がイベント発生を検知する毎、提示部12による関連情報の提示が更新される。

【0066】なお、蓄積部11によるデータベースの更新は、所定のタイミングにおいて実行される。データベースの更新処理は、図40のフローチャートを参照して後述する。また、蓄積部11によるデータベースの更新時には、記憶部49に記録した特徴ベクトルが、例えば、電子メールの送受信の頻度や回数、電子メールに含まれる単語の品詞の種類(特定の地域や名称を示す固有名称など)に応じて修正される。

【0067】図2は、エージェントプログラム1乃至ワープロプログラム3がインストールされて実行されるパーソナルコンピュータの構成図を示している。なお、当然ながら、本発明はパーソナルコンピュータの他、テレビジョン受像機、ホームサーバシステム、ハードディスクレコーダ、ゲーム機器、カーナビゲーションシステム、携帯電話、PDA等の情報電子機器において利用できる。

【0068】このパーソナルコンピュータは、CPU(Central Processing Unit)41を内蔵している。CPU41には、バス44を介して入出力インターフェース45が接続されている。入出力インターフェース45には、キーボード、マウスなどの入力デバイスよりなる入力部46、処理結果としての例えば音声信号を出力する出力部47、処理結果としての画像を表示するディスプレイなどよりなる表示部48、プログラムや構築されたデータベースなどを格納するハードディスクドライブなどよりなる記憶部49、インターネットに代表されるネットワークを介してデータを通信するLAN(Local Area Network)カードなどよりなる通信部50、および、磁気ディスク2、光ディスク3、光磁気ディスク4、または半導体メモリ5などの記録媒体に対してデータを読み書きするドライブ51が接続されている。バス44には、ROM(Read Only Memory)42およびRAM(Random Access Memory)43が接続されている。

【0069】本発明のエージェントプログラム1は、磁気ディスク2乃至半導体メモリ5に格納された状態でパーソナルコンピュータに供給され、ドライブ51によって読み出されて、または通信部50がネットワークを介して取得して、記憶部49に内蔵されるハードディスクドライブにインストールされている。記憶部49にインストールされているエージェントプログラム1は、

入力部46に入力されるユーザからのコマンドに対応するCPU41の指令によって、記憶部49からRAM43にロードされて実行される。なお、パーソナルコンピュータの起動時において自動的にエージェントプログラム1が実行されるように設定することも可能である。

【0070】また記憶部49に内蔵されるハードディスクドライブには、エージェントプログラム1の他、メール2、およびワープロプログラム3、WWW(World Wide Web)ブラウザなどのアプリケーションプログラムもインストールされており、エージェントプログラム1と同様に、入力部46に入力されるユーザからの起動コマンドに対応するCPU41の指令によって、記憶部49からRAM43にロードされて実行される。

【0071】次に、エージェントプログラム1によるデータベース作成処理について、図3のフローチャートを参照して説明する。このデータベース作成処理は、エージェントプログラム1が実行する処理のうちの1つであり、エージェントプログラム1が起動された状態において、データベースがまだ作成されていないときに開始される。

【0072】ステップS1において、文書取得部21は、データベース作成の素として分析する文書(例えば、エージェントプログラム1が実行される以前に送受信された電子メール。以下、分析対象電子メールと記述する)を、記憶部49に内蔵されるハードディスクドライブから選択的に取得して文書属性処理部22および文書内容処理部23に供給する。

【0073】ステップS1の処理、すなわち、分析対象電子メール選択処理の詳細については、図4を参照して説明する。

【0074】ステップS21において、文書取得部21は、ユーザが送信した電子メールが保存されている送信フォルダを参照し、直近の所定期間(例えば、最近の一週間)に送信した電子メールの数が所定数(例えば、100通)以上存在するか否かを判定する。直近の所定期間に送信した電子メールの数が所定数以上存在すると判定された場合、処理はステップS22に進む。ステップS22において、文書取得部21は、日時条件およびアドレス属性条件を設定する。

【0075】ステップS22の処理、すなわち、日時条件およびアドレス属性条件を設定処理の詳細については、図5を参照して説明する。ステップS31において、文書取得部21は、送信フォルダに存在する電子メールの数が所定数(例えば、1000通)以上であるか否かを判定する。

【0076】ステップS31において、送信フォルダに存在する電子メールの数が所定数以上であると判定された場合、処理はステップS32に進む。ステップS32において、文書取得部21は、分析対象電子メールを選択するための日時条件を「1年以前を除去」に設定す

る。ステップS33において、文書取得部21は、分析対象電子メールを選択するためのアドレス属性条件を「"To"以外を除去」に設定する。また、文書取得部21は、アドレス条件(アドレスリスト)を抽出する対象を送信フォルダに設定する。

【0077】反対に、ステップS31において、送信フォルダに存在する電子メールの数が所定数より少ないと判定された場合、処理はステップS34に進む。ステップS34において、文書取得部21は、日時条件を「3年以前を除去」に設定する。ステップS35において、文書取得部21は、アドレス属性条件を「"To"、"Cc"以外を除去」に設定する。また、文書取得部21は、アドレス条件を抽出する対象を送信フォルダおよび受信フォルダに設定する。

【0078】以上のような日時条件およびアドレス属性条件設定処理により、送信した電子メールの数に対応して、分析対象電子メールの日時条件とアドレス属性条件が設定された後、処理は図4のステップS23にリターンする。

【0079】なお、日時条件およびアドレス属性条件設定処理は、上述した2種類の選択だけでなく、例えば、送信フォルダのメール数に応じていくつかの区間を設け、それに応じて、日時条件を任意の年数で細かく区切ったり、受信簿に対するアドレス属性条件にさらにfrom, reply to等を加えた選択肢を増やすなどしてもよい。

【0080】ステップS23において、文書取得部21は、送信フォルダ(または受信フォルダ)に存在する電子メールを、ステップS22で設定した日時条件およびアドレス属性条件に基づいてフィルタリングすることにより、電子メールの数を絞り込む。ステップS24において、文書取得部21は、ステップS23でフィルタリングされた各電子メールの宛先(または送信元)をリスト化するとともに、各宛先の出現回数をカウントし、出現回数が多い上位n個のアドレスを判定して、アドレス条件を「上位n個のアドレスから送受信された電子メールを抽出」に設定する。

【0081】ステップS25において、文書取得部21は、全ての電子メール、すなわち、送信フォルダ、受信フォルダ、およびその他のフォルダの存在する電子メールのうち、ステップS22で設定した日時条件およびステップS24で設定したアドレス条件に基づいてフィルタリングすることにより、分析対象電子メールを選択する。

【0082】なお、ステップS21において、ユーザが送信した電子メールが保存されている送信フォルダを参照し、直近の所定期間に送信した電子メールの数が所定数より少ないと判定された場合、処理はステップS26に進む。ステップS26において、文書取得部21は、ユーザが送信した電子メールが保存されている受信

フォルダを参照し、直近の所定期間（例えば、最近の一週間）に受信した電子メールの数が所定数（例えば、100通）以上存在するか否かを判定する。直近の所定期間に受信した電子メールの数が所定数以上存在すると判定された場合、処理はステップS22に進み、それ以降の処理が繰り返される。

【0083】 反対に、ステップS26において、直近の所定期間に受信した電子メールの数が所定数よりも少ないと判定された場合、この段階でデータベース作成処理は終了される。

【0084】 以上のように分析対象電子メールが選択された後、処理は図3のステップS2にリターンする。

【0085】 ステップS2において、文書属性処理部22は、ステップS1の処理で文書取得部21から供給された分析対象電子メールから属性情報（メッセージID等のヘッダ情報）を抽出し、その属性情報に基づき、分析対象電子メールを話題毎に分類して（すなわち、話題毎にグループ化して）、話題毎に話題ファイルを生じして文書内容処理部23および文書特徴データベース作成部24に供給する。

【0086】 図6は、ステップS2において作成される話題ファイル61の一例を示している。話題ファイル61は、各話題ファイルを識別するためのトピックスID62、当該話題に属する最古の電子メールの通信時間を示す日時情報63、当該最古の電子メールの題名などを示すサブジェクト情報64、当該話題に属する電子メールの送信元または宛先の電子メールアドレスとなるメンバー情報65、当該話題に属する各電子メールを特定するメールメッセージID66、当該話題に属する電子メールの本文に含まれる単語から構成される単語ベクトル67、当該話題に属する電子メールの本文を連結した連結本文68、および当該話題に含まれる全ての単語の評価値から成る特徴ベクトル69から構成される。

【0087】 トピックスID62として、例えば当該話題に属する最古の電子メールの通信時間を用いるようにしてもよい。

【0088】 なお、連結本文68は、当該話題に属する電子メールのうち、送信フォルダやその他のフォルダに存在する電子メールの本文を連結するようにする。

【0089】 図7は、単語ベクトル67を構成する複数の単語70に含まれる要素を示している。すなわち、単語70には、当該単語自身の文字列71、当該単語の品詞（名詞の種類）72、当該話題における当該単語の頻度73、および当該話題における当該単語の評価値74を記録するための構成を有している。なお、単語70の各要素の中身はステップS2の処理段階では生成されず、以降の処理において生成される。

【0090】 また、特徴ベクトル69も、ステップS2の処理段階では生成されず、以降の処理において生成される。

【0091】 図3に戻る。ステップS3において、文書属性処理部22は、ステップS2で生成した話題を選択する。ステップS3の処理、すなわち第1次話題選択処理について、図8のフローチャートを参照して説明する。

【0092】 ステップS41において、文書属性処理部22は、ステップS2で生成した話題の数が所定数以上存在するか否かを判定する。生成した話題の数が所定数以上存在すると判定された場合、処理はステップS42に進む。ステップS42において、文書属性処理部22は、生成した話題を選択するための構成メール数条件を「a（例えば4）通以下を削除」に設定する。

【0093】 反対に、ステップS41において、生成した話題の数が所定数よりも少ないと判定された場合、処理はステップS43に進む。ステップS43において、文書属性処理部22は、生成した話題を選択するための構成メール数条件を「b（aよりも小さい数、例えば2）通以下を削除」に設定する。

【0094】 ステップS44において、文書属性処理部22は、上段の処理で設定した構成メール数条件に基づき、ステップS2で生成した話題をフィルタリングする。すなわち、例えば、上段の処理で構成メール数条件を「a通（例えば4通）以下を削除」に設定した場合、4通以下の電子メールから構成される話題を削除し、5通以上の電子メールから構成される話題だけを選択する。

【0095】 さらに、直近の所定期間（例えば、最近の一週間）に通信した電子メールを含まない話題を削除するようにしてもよい。

【0096】 このようにして第1次話題選択処理を実行した後、処理は図3のステップS4にリターンする。

【0097】 なお、第1次話題選択処理における構成メール数条件の設定は、上述した2種類の選択だけでなく、例えば、話題の数に応じていくつかの区間を設けて、その区間ごとに構成メール数条件を決定するようにしてもよい。

【0098】 ステップS4において、文書内容処理部23は、選択された各話題に対応する話題ファイル61の連結本文68に形態素解析を実行する。ステップS4における形態素解析処理の詳細について、図9のフローチャートを参照して説明する。

【0099】 ステップS51において、文書内容処理部23は、選択された各話題のうち、形態素解析を施していないものがあるか否かを判定する。形態素解析を施していないものが存在すると判定された場合、処理はステップS52に進む。ステップS52において、文書内容処理部23は、形態素解析を施していない話題を1

つ選択し、対応する話題ファイル61の連結本文68を読み出して形態素解析を施し、連結本文68に含まれる単語を抽出する。

【0100】このように、話題ファイル61の連結本文68に対して形態素解析を施す処理は、話題ファイル61を構成する電子メールの各本文に対して形態素解析を施す処理に比較して、処理する文章は長くなるが処理回数が1回で済むので、処理に要するリソースの浪費を抑止することができる。

【0101】ステップS53において、文書内容処理部23は、ステップS52で抽出した単語のうち、品詞が名詞（一般名詞、変換接続名詞、地名、人名、興味がある用語を含む）であるものを抽出する。ステップS54において、文書内容処理部23は、抽出した名詞である単語を並べ、当該話題に対応する単語ベクトル67を生成する。

【0102】ステップS55において、文書内容処理部23は、話題単語テーブル81（図10）にステップS54で生成した単語ベクトル67に対応する記録を追加するとともに、ステップS54で生成した単語ベクトル67を構成する単語の記録を、話題評価値テーブル93を含む単語インデックステーブル91（図11）に追加する。なお、話題単語テーブル81、単語インデックステーブル91、および話題評価値テーブル93は、いずれもハッシュテーブル(hash table)である。

【0103】図10は、話題単語テーブル81の構成例を示している。話題単語テーブル81は、各話題に対するトピックID62と、それに対応する単語ベクトル67が記録されており、トピックID62を入力として、対応する単語ベクトル67を出力する。

【0104】図11は、単語インデックステーブル91の構成例を示している。単語インデックステーブル91は、各単語ベクトル67を構成する単語名92と、それに対応する話題評価値テーブル93の組が複数記録されており、単語名92を入力として、話題評価値テーブル93を出力する。

【0105】図12は、話題評価値テーブル93の構成例を示している。話題評価値テーブル93は、単語名92に対応する単語が含まれる話題のトピックID101と、当該話題における当該単語の評価値102が記録されており、トピックID101を入力として、当該話題における当該単語の評価値102を出力する。

【0106】このような構成の話題単語テーブル81乃至話題評価値テーブル93を生成することにより、トピックID62および単語名92のどちらか一方を入力としても、対応する他方を容易に検索することが可能となる。

【0107】この後、処理はステップS51に戻り、以降の処理が繰り返される。その後、ステップS51において、選択された各話題のうち、形態素解析を施してい

ないものが存在しないと判定された場合、形態素解析処理は終了され、処理は図3のステップS5にリターンする。

【0108】ステップS5において、文書内容処理部23は、以降における処理を軽減するために、これまでの処理で抽出された単語、すなわち、各話題にそれぞれ対応する単語ベクトルに含まれる単語のうち、話題の内容に関連が薄いと考えられる単語、あいさつなどの日常的な単語等（以下、不要語と記述する）を除去する。

【0109】ステップS5における不要語削除処理について、図13のフローチャートを参照して説明する。ステップS61において、文書内容処理部23は、単語ベクトルが小さい話題、すなわち、対応する単語ベクトルを構成する単語の数が所定数（例えば、5個）以下である話題を除去する。

【0110】ステップS62において、文書内容処理部23は、ステップS4の処理で生成した単語インデックステーブル91に記録されている単語のうち、以降の処理の対象としていない単語が存在するか否かを判定する。処理対象としていない単語が存在すると判定された場合、処理はステップS63に進む。ステップS63において、文書内容処理部23は、単語インデックステーブル91に記録されている、処理対象としていない単語のうちの1つを処理対象の単語に選択する。

【0111】ステップS64において、文書内容処理部23は、処理対象の単語を入力として、単語インデックステーブル91を参照することにより、対応する話題評価値テーブル93を取得し、取得した話題評価値テーブル93に記録されているトピックID101の数をカウントすることによって、処理対象の単語を含む話題の数を取得する。

【0112】ステップS65において、文書内容処理部23は、処理対象の単語を含む話題の数が所定数以上であるか否かを判定する。処理対象の単語を含む話題の数が所定数以上であると判定された場合、処理はステップS66に進む。ステップS66において、文書内容処理部23は、処理対象の単語を、不要語ベクトル（不要語を構成要素とする）に追加する。これにより、多数の話題に共通して含まれると考えられるあいさつなどの日常的な単語が不要語ベクトルに追加される。

【0113】ステップS67において、文書内容処理部23は、不要語である処理対象の単語に対応する記録を削除するため、各話題にそれぞれ対応する話題ファイル61、話題単語テーブル81、単語インデックステーブル91、および話題評価値テーブル93を更新する。この後、処理はステップS62に戻り、以降の処理が繰り返される。

【0114】なお、ステップS65において、処理対象の単語を含む話題の数が所定数よりも小さいと判定された場合にも、ステップS66およびS67はスキップさ

れて、処理はステップS62に戻る。

【0115】その後、ステップS62において、ステップS4の処理で生成した単語インデックステーブル91に記録されている単語のうち、以降の処理の対象としていない単語が存在しないと判定された場合、処理はステップS68に進む。ステップS68において、文書内容処理部23は、再びステップS61の処理と同様に、単語ベクトルが小さい話題、すなわち、対応する単語ベクトル67を構成する単語の数が所定数（例えば、5個）以下である話題を除去する。これにより、日常的な単語ばかりで構成されているとみなされる話題が除去される。この段階で、話題は特徴的な単語から構成される単語ベクトル67によって集算されることになる。処理は図3のステップS6に戻る。

【0116】ステップS6において、文書内容処理部23は、不要語が削除された各単語ベクトル67を構成する全ての単語について、その出現頻度および複数の文書に亘る分布状況を求め、各話題における評価値を演算する。評価値の演算は、例えば、idf法を用いる。ステップS7において、文書特徴データベース作成部24は、ステップS6で演算した各単語に対する評価値を、次の条件に基づいて修正する。

【0117】例えば、送信した電子メールに含まれる単語の評価値がより大きくなるように修正を行う。送信した電子メールに含まれる単語を特定するためには、ステップS2の処理で生成した各話題に対応する話題ファイル61の連結本文68に挿入した、所定の文字列（例えば“soshin-shuryo”）を検出し、当該所定の文字列以前の単語を、送信した電子メールに含まれる単語として特定すればよい。

【0118】また例えば、属する電子メールの数が多い話題に含まれる単語の評価値が、属する電子メールの数に対応して大きくなるように修正を行う。例えば、属する電子メールの数を m とした場合、修正前の評価値に対し、 $1/\log(\text{次数} \times a \cdot m)$ （ a は定数）、対数回数 $\log(m)$ などの単調増加関数乗算する。この修正は、電子メールのような時間的に継続するやりとりでは、以前の文書に登場した単語が、次の文書では指示代名詞によって置換されることが多いので、話題に属する電子メールの数が多くなると、単語の評価値が相対的に小さくなってしまいう傾向にあることを考慮したものである。

【0119】さらに例えば、通信頻度が高い相手と通信した電子メールに含まれる単語、および特定名詞（定義した興味語、一般名、地名、組織名など）などの評価値がより大きくなるように修正を行う。なお、特定名詞に対する評価値の修正方法については、特開2001-79511号として提案した発明を適用することができる。

【0120】ステップS8において、文書特徴データベース作成部24は、ステップS6で演算され、ステップ

S7で修正された各単語に対する評価値を、話題ファイル61および話題単語テーブル81の単語ベクトル67、並びに単語インデックステーブル91の中の話題評価値テーブル93に記録する。これにより、各単語ベクトル67を構成する単語70の全ての要素が決定されたことになる。また、文書特徴データベース作成部24は、各話題にそれぞれ対応する特徴ベクトル69を確定して記録する。さらに、文書特徴データベース作成部24は、各単語ベクトル67について、構成する単語を評価値が大きい順に並べ替える。

【0121】ステップS9において、文書特徴データベース作成部24は、この段階で残っている話題を再び選択する。ステップS9の処理、すなわち第2次話題選択処理について、図14のフローチャートを参照して説明する。なお、この第2次話題選択処理は、各話題に対して実行される。

【0122】ステップS71において、文書特徴データベース作成部24は、話題に対応する単語ベクトル67を構成する単語のうち、評価値が最大のもの（あるいは、上位の2、3語）を検出する。ステップS72において、文書特徴データベース作成部24は、ステップS71で検出した単語の評価値が所定値以上であるか否かを判定する。検出した単語の評価値が所定値以上であると判定された場合、処理はステップS73に進む。

【0123】ステップS73において、文書特徴データベース作成部24は、当該話題に属する電子メールの最新の通信日時が直近の所定期間（例えば、最近1週間）以前であるか否かを判定する。最新の通信日時が直近の所定期間以前ではないと判定された場合、処理はステップS74に進む。ステップS74において、文書特徴データベース作成部24は、当該話題の最も評価値が高い単語を最近語ベクトル73に追加する。ステップS75において、文書特徴データベース作成部24は、当該話題を削除する。ステップS73乃至ステップS75の処理により、新しい話題が削除されるので、後述する関連情報の推薦に意外性を増やすことができる。

【0124】なお、ステップS72において、ステップS71で検出した単語の評価値が所定値よりも小さいと判定された場合、ステップS73およびステップS74はスキップされ、処理はステップS75に進む。

【0125】また、ステップS73において、当該話題に属する電子メールの最新の通信日時が直近の所定期間以前であると判定された場合、当該話題に対する第2次話題選択処理は終了され、次の話題に対する第2次話題選択処理が開始される。

【0126】その後、全ての話題に対して第2次話題選択処理を施した後、選択された話題のうち、対応する単語ベクトル73の上位に（すなわち、評価値が高い方の2、3番目まで）、最新語ベクトルに含まれる単語を含んでいるものを削除するようする。これにより、後述

する関連情報の推薦に意外性をより増やすことができる。処理は、図3のステップS10にリターンする。

【0127】ステップS10において、文書特徴データベース作成部24は、この段階で選抜されている話題にそれぞれ対応する各単語ベクトル67について、構成する単語の評価値の最大値に注目し、評価値の最大値が大きい順に所定数（例えば、200）だけ単語ベクトル67を抽出し、それぞれに対応する所定数の話題を推薦話題候補に確定する。

【0128】ステップS11において、文書特徴データベース作成部24は、ステップS10で確定した推薦話題候補に基づいて、推薦話題を確定する。ステップS11における推薦話題確定処理について、図15のフローチャートを参照して説明する。

【0129】ステップS81において、文書取得部21は、メール2の送信フォルダおよび受信フォルダから最近の所定期間（例えば、直近の1週間）に送受信した電子メールのうち、アドレス条件に合うものを取得する。なお、ここで取得された各電子メールは既にいずれかの話題に分類されている。

【0130】ステップS82において、文書属性処理部22は、既に生成されている全ての話題ファイル61のメールメッセージID66を参照することによって、ステップS81で取得した各電子メールが属する話題を特定する。

【0131】ステップS83において、文書特徴データベース作成部24は、ステップS82で特定された、最近の各話題にそれぞれ対応する特徴ベクトル69（以下、特徴ベクトルVcと記述する）を取得する。ステップS84において、文書特徴データベース作成部24は、各特徴ベクトルVcに対する、ステップS10で確定した推薦話題候補にそれぞれ対応する特徴ベクトル69（以下、特徴ベクトルVtと記述する）の類似性を判定するために、特徴ベクトルVcと特徴ベクトルVtとの全ての組み合わせの内積Sim(Vc, Vt)を次のように演算する。

$$\text{Sim}(Vc, Vt) = \frac{Vc \cdot Vt}{\sqrt{Vc \cdot Vc} \cdot \sqrt{Vt \cdot Vt}}$$

【0132】ここで、内積Sim(Vc, Vt)は、各特徴ベクトルVcに対する特徴ベクトルVtの類似性を判定するために用いるので、特徴ベクトルVcの絶対値|Vc|で除算する演算を省略することが可能となる。

【0133】ステップS85において、文書特徴データベース作成部24は、各特徴ベクトルVcに対して、内積演算結果が最大である特徴ベクトルVtを判別して、それに対応する推薦話題候補を推薦話題に確定する。この段階で、最新の電子メールのうち、アドレス条件にあったメールが属する話題の数と同数の推薦話題が確定される。

【0134】ステップS86において、文書特徴データベース作成部24は、ステップS85で確定した推薦話題の数が所定数（例えば、30）よりも少ないか否かを判定する。確定した推薦話題の数が所定数よりも少ないと判定された場合、処理はステップS87に進む。ステップS87において、文書特徴データベース作成部24は、ステップS85で確定した推薦話題の数が所定数に対して不足する分だけ、この段階で推薦話題に確定されていない推薦話題候補のうち、含まれる単語の評価値の最大値が高い話題から順番に推薦話題に追加する。

【0135】なお、ステップS86において、ステップS85で確定した推薦話題の数が所定数以上であると判定された場合、ステップS87の処理はスキップされる。

【0136】このようにして、所定数だけ推薦話題が確定された後、処理は図3のステップS12にリターンする。

【0137】ステップS12において、関連情報検索部25は、ステップS11で確定された推薦話題に対応する関連情報を、インターネット上のWebサイトを用いて検索する。ステップS12におけるWeb検索処理について、図16のフローチャートを参照して説明する。

【0138】ステップS91において、文書特徴データベース作成部24は、ステップS11で確定した推薦話題のうち、Web検索の対象としていない推薦話題が存在するか否かを判定する。Web検索の対象としていない推薦話題が存在すると判定された場合、処理はステップS92に進む。ステップS92において、文書特徴データベース作成部24は、Web検索の対象としていない推薦話題の1つを選択する。

【0139】ステップS93において、文書特徴データベース作成部24は、選択した推薦話題に対応する特徴ベクトル69（または単語ベクトル67）を読み出し、その特徴ベクトル69を構成する単語のうち、評価値が上位側の2単語（1単語、あるいは3単語以上でもよい）を取得して連結し、検索語として関連情報検索部25に供給する。

【0140】ステップS94において、関連情報検索部25は、インターネット上の検索エンジンにアクセスし、文書特徴データベース作成部24から供給された検索語を送信する。ステップS95において、関連情報検索部25は、検索エンジンから検索結果としてのWebページのタイトルとURLを取得する。

【0141】ステップS96において、関連情報検索部25は、取得した検索結果を、予め設定された特定単語に基づいてフィルタリングする。具体例には、他人が見ても興味を持たないような一般性が低いWebページのタイトルに含まれると思われる特定単語（日記、議事録、予定、行事、ミーティング等）がWebページのタイトルに含まれる検索結果を除外する。この後、関連情報検索

部24は、残った検索結果（WebページのタイトルとURL）に関連情報として文書特徴データベース作成部24に供給する。

【0142】処理はステップS91に戻り、以降の処理が繰り返される。その後、ステップS91において、ステップS11で確定した推薦話題のうち、Web検索の対象としていない推薦話題が存在しないか判定された場合、処理はステップS97に進む。

【0143】ステップS97において、文書特徴データベース作成部24は、予め設定されている作り込み推薦用単語組（例えば、（旅行、温泉）、（観光、ホテル）、（グルメ、レストラン）、（スポーツ、サッカー）、（ソニー、新製品）等）のうち、Web検索の対象としていない作り込み推薦用単語組が存在するか否かを判定する。なお、作り込み推薦用単語組は、ユーザが任意に追加、または削除することができる。

【0144】Web検索の対象としていない作り込み推薦用単語組が存在すると判定された場合、処理はステップS98に進む。ステップS98において、文書特徴データベース作成部24は、Web検索の対象としていない作り込み推薦用単語組の1つを選択する。処理はステップS94に進み、以降の処理が繰り返される。

【0145】その後、ステップS97において、予め設定されている作り込み推薦用単語組のうち、Web検索の対象としていない作り込み推薦用単語組が存在しないと判定された場合、Web検索処理を終了して、処理は図3のステップS13にリターンする。

【0146】ステップS13において、文書特徴データベース作成部24は、関連情報検索部25から供給された関連情報を、検索語に対応付けて記憶部49に記録することにより、データベースを作成する。なお、ステップS12以降の処理は、ステップS11までの一連の処理に継続して実行される場合と、一連の処理に継続せず、所定のタイミングにおいて実行される場合がある。

【0147】以上のデータベース作成処理が実行されることにより、送受信した電子メールの文書に対応した関連情報がデータベース内に蓄積されることになる。なお、データベース作成処理は、エージェントプログラム1が実行された場合に開始されるものとしたが、任意のタイミングで開始させることも可能である。さらに、このようにして作成されたデータベースは、所定の条件が満たされたときに更新される（更新のタイミングについては、図4を参照して後述する）。

【0148】また、データベース作成処理をユーザが強制的に中断することができるようにするために、中断要求があった場合、中断された時点で整理済みの文書を記録し、再開要求があった場合、未処理の文書から処理を再開するようにしてもよい。

【0149】次に、エージェントプログラム1による関連情報提示処理について、図17のフローチャートを参

照して説明する。この関連情報提示処理は、上述したデータベース作成処理とは異なり、エージェントプログラム1が実行されている間、繰り返し実行される。

【0150】ステップS111において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、エージェントプログラム1の終了が指示されたか否かを判定し、エージェントプログラム1の終了が指示されていないと判定した場合、ステップS112に進む。

【0151】ステップS112において、イベント管理部31は、イベント発生（メーラ2の電子メールの送受信の完了等）を監視し、イベント発生が検知されない場合、ステップS111に戻り、上述した処理が繰り返して実行される。

【0152】ステップS112において、イベント発生が検知された場合（例えば、新たな電子メールの送受信が検知された場合）、処理はステップS113に進む。ステップS113において、イベント管理部31は、イベント発生をデータベース問い合わせ部32に通知する。データベース問い合わせ部32は、イベント管理部31からのイベント発生の通知に対応して、イベント発生に対応する文書（送受信された電子メール）を取得し、その文書の形態素解析を施して、不要語を除外した単語（特長語）を抽出し、各単語の評価値を演算する。これにより、イベント発生に対応する文書（いまの場合、電子メール）の特長ベクトルが算出される。

【0153】ステップS114において、データベース問い合わせ部32は、文書特徴データベース作成部24が作成したデータベースを検索し、ステップS113の処理で算出された特長ベクトルと、データベースに記録されている話題毎の特長ベクトルとの内積を両者の類似度として算出し、類似度が所定の条件（例えば、類似度が最大、もしくは類似度が所定の閾値以上）を満たす話題を抽出する。

【0154】ステップS115において、データベース問い合わせ部32は、ステップS114の処理で抽出された話題に含まれる各単語のうち、評価値の時系列推移に着目して、以下で説明する条件1および条件2を満たす単語（重要語）を選択する。さらに、データベース問い合わせ部32は、このようにして選択した単語（重要語）に対応する関連情報と、イベント管理部31を介して、または直接的に、関連情報提示部33に供給する。

【0155】ここで、単語の選択条件について、図18を参照して説明する。図18は、データベースに蓄積されている単語の評価値の時系列推移の例を示している。

【0156】例えば、条件1を「単語の評価値が、現時点以前の所定の期間X（例えば、2週間）、所定の閾値A以下であること」とする。また例えば、条件2を「現時点以前の所定の期間Y（例えば、5週間）において、異なる2以上の話題で、単語の評価値が所定の閾値B以

上であること」とする。なお、条件3として、「条件2における異なる2以上の話題のうち、最も古い話題と最も新しい話題が所定の期間Z以上離れていること」を追加すればさらに好ましい。

【0157】このような条件を用いることにより、ユーザが高い関心を持っていると思われる単語(重要語)を選択することが可能となる。特に、条件1を設けることにより、現時点に近い話題に含まれる単語は除外されるので、ユーザが現時点で意識して意外性がないと思われる関連情報(新しすぎる情報)が選択されることを避けることができ、かつ、かなり以前の話題に含まれる単語も除外されるので、ユーザが現時点で思い出すことができないと思われる関連情報(古すぎる情報)が選択されることも避けることができる。

【0158】図17の説明に戻る。この段階までに、イベント発生(いまの場合、電子メールが送受信されたこと)に対応する関連情報が選択されることになるが、ステップS112において、例えば、メーラ2がアクティブとされたことがイベント発生として検知された場合には、上述したデータベース作成処理によって確定された、推薦する関連情報が用いられる。このとき、重要語がデスクトップに表示される。

【0159】ステップS116において、エージェント制御部13は、ステップS115の処理で選択した単語が含まれている文書の属性情報を、選択(推薦)した理由としてデスクトップに表示させるとともに、対応する関連情報を表示するかをユーザに入力ウィンドウ181(図26)をデスクトップに表示させる。

【0160】なお、話題は、グループ化された1以上の文書から構成されるので、重要語が含まれる文書も複数存在する場合がある(すなわち、重要語が含まれている文書の属性情報が複数存在する場合がある)。そこで、例えば、重要語が含まれている文書のうち、最もまたは最新の文書の属性情報を表示させるようにするか、または、任意に指定された文書の属性情報を表示させるようにする。また、入力ウィンドウ181を表示させずに、直接、デスクトップ上に関連情報を表示させるようにしてもよい。

【0161】ステップS117において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、ステップS116の処理で表示された入力ウィンドウ181に呼応して、ユーザが入力ウィンドウ181の「見る」ボタンを選択したか否かを判定する。ステップS117において、ユーザが「見る」ボタンを選択したと判定された場合、ステップS118に進む。なお、入力ウィンドウ181には、「見る」ボタンおよび「見ない」ボタン以外に他の情報を表示したりすることができる。あるいは、表示しないようにすることもできる。

【0162】ステップS118において、関連情報提示

部33は、イベント管理部31を介してデータベース問い合わせ部32から供給された関連情報をデスクトップに表示させる。この関連情報は、1または複数同時に提示することができる。

【0163】なお、関連情報として表示される情報は、キーワードが付与された所定のデータベースに蓄積された情報であれば、Webページのタイトルでなくてもかまわない。例えば、所定のデータベースに蓄積されている情報のインデックスを表示するようにして、ユーザのアクセス指令に対応して、そのインデックスのさらに詳細な情報を表示させるようにしてもよい。

【0164】ステップS119において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、ステップS118の処理により関連情報として表示されたWebページのタイトルに対して、ユーザがアクセスを指令したと判定した場合、ステップS120に進む。ステップS120において、WWWブラウザが起動され、対応するWebページに対するアクセスが開始される。

【0165】ステップS119において、ステップS118の処理により関連情報として表示されたWebページのタイトルに対して、ユーザが記録を指令したと判定された場合、ステップS121に進む。ステップS121において、エージェントプログラム1は、対応するWebページのタイトルおよびURLを、提示履歴を表示するスクラップ帳ウィンドウ174(図21)に記録する。

【0166】ステップS119において、ステップS118の処理により関連情報として表示されたWebページのタイトルに対して、ユーザからの指令もなされずに所定の時間が経過したと判定された場合、ステップS120またはステップS121の処理はスキップされて、ステップS111に戻り、上述した処理が繰り返して実行される。

【0167】なお、ステップS117において、ユーザが「見る」ボタンを選択しないか判定された場合、ステップS118乃至ステップS121の処理はスキップされて、ステップS111に戻り、上述した処理が繰り返して実行される。さらに、ステップS111において、ユーザによりエージェントプログラム1の終了が指示されたと判定された場合、関連情報提示処理は終了される。

【0168】ここで、関連情報提示処理に関して、イベント発生に対応する電子メールを効率よく取得する方法について説明する。

【0169】まず、メーラ1として適用する既存の多数の電子メール送受信用ソフトウェアが電子メールの保持形式に関し、次のような1つの特徴を有していることに着目する。

【0170】第1の特徴は、メーラにおける1つのフォルダは、パーソナルコンピュータにおける1つの電子メールボックスファイルに対応していることである。

【0171】第2の特徴は、新たに受信した電子メールは、特定のフォルダに格納されるようになっており、パーソナルコンピュータでは当該フォルダに対応するファイルの末尾に追加され、このとき、1つのファイルには一般に複数の電子メールの文章が含まれるので、各電子メールの文章の境界に、特定の文字列パターン（メーラによって異なる）からなる行が挿入されていることである。

【0172】第3の特徴は、送信した電子メールの記録も、同様の形式で特定のファイルに保存されることである。

【0173】第4の特徴は、送受信した電子メールが含まれるファイルはサイズが比較的大きい（数百キロバイト乃至1キロバイト）ことである。

【0174】以上の第1乃至第4の特徴を考慮して、次の手順によってイベント発生に対応する電子メールを取得する。始めに、電子メールボックスファイルの更新日時を検出し、新たな電子メールが追加されたか否かを判断する。次に、新たに電子メールが追加された電子メールボックスファイルを末尾から先頭方向に1行ずつ操作して、各電子メールの文章の境界を示す特定の文字列を検出する。境界を示す文字列を検出した場合、その位置から電子メールボックスファイルの末尾までデータを抽出する。

【0175】このような手順により、イベント発生に対応する電子メールを効率的に取得することが可能となる。

【0176】次に、上述した関連情報提示処理に関し、同一の電子メールに対して何度も関連情報を提示しないようにする手法について説明する。まず、関連情報を提示した電子メールのメッセージIDを記録するためのデータ構造を設定する。そして、イベントが発生した場合、そのイベントに対応する電子メールのメッセージIDを取得して、設定したデータ構造と比較する。データ構造の中に同じメッセージIDが存在する場合、その電子メールに対しては既に関連情報を提示しているので、関連情報を提示しないようにする。一方、データ構造の中に同じメッセージIDが存在しない場合、その電子メールに対しては関連情報を提示し、メッセージIDをデータ構造に記録する。

【0177】このような手法を用いることにより、同一の電子メールに対して何度も関連情報を提示するような事態の発生を抑制することが可能となる。

【0178】次に、上述した関連情報提示処理に関連する、主にエージェントの動作および台詞等について、図19および図20のフローチャートを参照して、詳細に説明する。

【0179】例えば、エージェントプログラム1が起動されている状態においてメーラ2が起動された場合、ステップS131において、エージェント制御部13は、

例えば、図21に示されるように、メーラ2のウィンドウ（以下、メーラウィンドウと記載する）171の表示と重畳しない位置に、エージェント172を登場させる。

【0180】なお、エージェント172の登場は、例えば、図22A乃至図22Dに示す画像が順次表示されることによって、エージェント172が前転しながらデスクトップ上に出現する動画が表現される。エージェント172の登場とともに、エージェント172の台詞として吹き出し173と、保存されている関連情報が一覧表示されているスクラップ帳ウィンドウ174（後述）が表示される。吹き出し173の中には、例えば図23に示されるように、登場の挨拶「おはよう、SAITOさん！」と、自己紹介「ぼく、aiだよ。」の台詞が表示される。

【0181】また、吹き出し173の表示と同期して、吹き出し173に表示された台詞と同じ意味を持つ他の言語（例えば、英語の場合、「Good morning, SAITO」、「I'm AI!」）の音声信号が音声合成部（図示せず）によって合成されて出力するようにすることができる。なお、吹き出し173に表示された言語（いまの場合、日本語）と音声信号の言語（いまの場合、英語）を同じ言語に統一してもよい。なお、以降に表示される吹き出し173にも対応する音声信号が同期して出力されるように設定できる。

【0182】ただし、吹き出し173の表示の有無や台詞に対応する音声の出力の有無はエージェントプログラム1が適宜設定するか、ユーザが任意に設定できるようにすることが可能である。

【0183】その後、エージェント172の表示は、ステップS132において、例えば図24に示されるように、待機中の様子（手を顔に組み、つま先を上下させる）を示す動画に推移される。

【0184】ステップS133において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドに応じて、メーラ2が終了されたか否かを判定する。メーラ2が終了されていないと判定された場合、処理はステップS134に進む。

【0185】ステップS134において（上述した図17のステップS112に対応する）、メーラ2は、ユーザから何らかのコマンド（電子メールの送受信、電子メールの編集、あるいは関連情報の編集等）が入力されたか否かを判定し、何らかのコマンドが入力されたと判定した場合、ステップS135に進み、コマンドに対応する処理を開始する。

【0186】ステップS135において、エージェントプログラム1のイベント管理部31は、電子メールの送信、受信、または編集のコマンドが入力されたか否かを判定する。電子メールの送受信または編集のコマンドが入力されたと判定された場合、処理はステップS136

に進む。

【0187】ステップS136において、エージェント制御部13は、エージェント172の表示を、図24に示した特機中の様子から、例えば図25に示されるように、作業中の様子（手足を激しく移動する）を示す動画に推移させる。この期間に、図17のステップS113乃至S115の処理（ユーザーに推薦する関連情報を選択する処理）が実行される。

【0188】ステップS137において、エージェントプログラム11は、コマンドに対応して開始されたメーラ2の処理（例えば、電子メール送信など）が継続中であるかを判定し、メーラ2の作業中の処理が終了するまで判定処理を繰り返し実行する。すなわち、メーラ2の作業中の処理が終了するまで、エージェント制御部13は、エージェント172の表示を、図25に示した作業中の状態のまま待機する。

【0189】ステップS137において、メーラ2の処理が継続してない、すなわち、コマンドに対応して開始されたメーラ2の作業中の処理が終了したと判定された場合、処理はステップS138に進む。

【0190】ステップS138において、エージェントプログラム11は、再度、入力部46に入力されたユーザーからのコマンドに応じて、メーラ2が終了されたか否かを判定する。メーラ2が終了されたらいいと判定された場合、処理はステップS139に進む。

【0191】ステップS139（図12のステップS116に対応する）において、エージェント制御部13は、ステップS137のメーラ2の処理が電子メール送信であった場合、エージェント172の吹き出し73に、例えば、台詞「今、Aさんにメール送ったけど、某月某日にAさんと（タイトル）について話していたよね。その中でできた（キーワード）について、関連するページを見つけたよ。見てみる？」と表示させる。

【0192】また、ステップS137のメーラ2の処理が電子メール受信であった場合には、例えば台詞「今、Aさんからメールがきたけど、某月某日にAさんと（タイトル）について話していたよね。その中でできた（キーワード）について、関連するページを見つけたよ。見てみる？」と表示させる。

【0193】さらに、ステップS137のメーラ2の処理が電子メールの編集であった場合には、例えば台詞「今、Aさんにメールを書いているけど、某月某日にAさんと（タイトル）について話していたよね。その中でできた（キーワード）について、関連するページを見つけたよ。見てみる？」と表示させる。となる。

【0194】なお、表示される台詞のうち、「某月某日にAさんと（タイトル）について話していたよね。」の部分は、関連情報が選択（推薦）された理由に相当するが、この関連情報の選択理由の表示を、ステップS139において実行せずに、後述するステップS142の処

理（関連情報の表示）の後に表示するようにしてもよい。また、関連情報の選択理由の表示をユーザーの指示により任意のタイミング（例えば、メニューで理由を聞くコマンドを用意するなど）で実行するようにしてもよい。

【0195】また、タイマ31Aによる一定時間経過時の提示に関しては、「今、Aさんからメールがきたけど」等の特定イベントを示すような表現ではなく、例えば台詞の一部「某月某日にAさんと（タイトル）」について話していたよね。」だけを表示するようにする。

【0196】さらに、これらの吹き出し73は、関連情報を表示する前に提示してもよいし、あるいは、表示した後に提示してもよい。

【0197】吹き出し73に隣接する位置には、例えば図26に示されるように、入力ウィンドウ181が表示される。入力ウィンドウ181は、図26に示されるように、関連情報の表示を指示するときに選択する「見る」ボタン、関連情報を表示させない時に選択する「見ない」ボタン、関連情報が選択された背景（選択理由）の再表示を指示するときに選択する「背景をもう一度教えて」ボタンが表示される。

【0198】入力ウィンドウ181が表示された状態で、ステップS140において、エージェント制御部13は、エージェント172の表示を、図26に示した特機中の様子を示す動画に推移させる。ステップS141（図17のステップS117に対応する）において、エージェントプログラム11は、入力ウィンドウ181の中の「見る」ボタン、「見ない」ボタン、または「背景をもう一度教えて」ボタンのいずれかがユーザーにより選択されたか否かを判定する。このウィンドウは表示しなくてもよい。

【0199】ステップS141において、入力ウィンドウ181の「見る」ボタンが選択されたらと判定された場合、処理はステップS142に進む。ステップS142（図17のステップS118に対応する）において、エージェント制御部13は、例えば、図28および図29に示されるように、関連情報として推薦URL191を表示させ、エージェント172の表示を、表示された推薦URL191を指し示す動画に推移させるとともに、吹き出し73に、台詞「どう？」を表示させる。推薦URL191には、通常、推薦されるWebページのタイトルが表示され、推薦URL191の上にマウスカーソルが置かれたときだけURLも重畳して表示される。推薦URL191は、マウスカーソルでドラッグすることにより移動可能である。

【0200】ステップS143（図17のステップS119に対応する）において、エージェントプログラム11は、表示した推薦URL191に対するユーザーのコマンドを検出する。表示される推薦URL191に対するユーザーのコマンドとしては、記録、アクセス、または消去等が

ある。

【0201】推薦URL 191に対する記録コマンドは、例えば、記録する推薦URL 191をスクラップ帳ウィンドウ 174までドラッグアンドドロップする方法、マウスの右ボタンをクリックし、表示されるメニューの中から記録を選択する方法などが考えられる。あるいは、推薦URLはすべて自動的に記録されるようにしてもよい。アクセスコマンドや消去コマンドについても同様に、Webブラウザのアイコンやゴミ箱のアイコンにドラッグアンドドロップする方法、マウスで右クリックし、表示されるメニューの中から選択する方法、あるいはクリックパブルにする方法などが考えられる。

【0202】ステップS143において、推薦URL 191に対する記録コマンドが検出された場合、ステップS144（図17のステップS121に対応する）において、エージェント制御部13は、エージェント172の表示を、例えば図30に示されるように、顔く動作に推移させる。スクラップ帳ウィンドウ 174の中には、記録が指示された推薦URL 191に対応するWebページのタイトルが追加表示される。

【0203】また、ステップS143で、推薦URL 191に対するアクセスコマンドが検出された場合、ステップS144（図17のステップS120に対応する）において、エージェント制御部13は、エージェント172の表示を、例えば図31Aおよび図31Bに示されるように、笑顔で喜ぶ様子に推移させる。吹き出し173には、台詞「わーい」が表示され、対応する音声信号が出力される。

【0204】また、ステップS143で、推薦URL 191に対する消去コマンドが検出された場合、ステップS144において、エージェント制御部13は、エージェント172の表示を、例えば図32Aおよび図32Bに示されるように、泣き顔で悲しみ失望した様子に推移させる。吹き出し173には、台詞「だめかぁ」が表示され、対応する音声信号が出力される。

【0205】この後、処理はステップS132に戻り、それ以降の処理が繰り返して実行される。

【0206】なお、ステップS141において、入力ウィンドウ 181の「見ない」ボタンが選択されたと判定された場合、処理はステップS32に戻り、それ以降の処理が繰り返して実行される。また、ステップS141において、入力ウィンドウ 181の「背景をもう一度教えて」ボタンが選択されたと判定された場合、処理はステップS139に戻り、ステップS139乃至S141の処理が繰り返される。

【0207】ステップS138において、メーラ2が終了されたと判定された場合、処理はステップS145に進む。ステップS145において、エージェント制御部13は、吹き出し173に、終了を惜しむ台詞「え、そんなぁ」を表示させ、対応する音声信号を出力させた

後、ステップS46において、エージェント72の表示を消失させる（図25を参照して後述する）。

【0208】ステップS135において、関連情報の編集を指示するコマンドが入力されたと判定された場合、処理はステップS147に進む。ステップS147において、関連情報提示部33は、関連情報編集用ウィンドウ（図示せず）を表示させ、エージェント制御部13は、エージェント172の表示を、図30に示した待機中の様子から、図29と同様に、関連情報編集用ウィンドウを指し示す様子に推移させる。その後、ユーザが関連情報編集用ウィンドウに対して編集のための入力を開始すると、ステップS148において、エージェント制御部13は、エージェント172の表示を、関連情報編集用ウィンドウを指し示す様子から、図25に示した作業中の様子を示す動きに推移させる。

【0209】ステップS149において、エージェントプログラム1は、関連情報編集処理が継続中であるか否かを判定し、関連情報編集処理が終了するまで判定処理を繰り返して実行する。すなわち、関連情報編集処理が終了するまで、エージェント制御部13は、エージェント172の表示を、図25に示した作業中の状態のまま待機する。

【0210】ステップS149において、関連情報編集処理が継続中ではない、すなわち、コマンドに対応して開始された関連情報編集処理が終了したと判定された場合、処理はステップS150に進む。

【0211】ステップS150において、エージェント制御部13は、エージェント172の表示を、図30と同様に、顔く様子に推移させる。吹き出し173には、台詞「変更したよ」と表示され、対応する音声信号が出力される。この後、処理はステップS132に戻り、それ以降の処理が繰り返して実行される。

【0212】ステップS134において、メーラ2に対してユーザから何らかのコマンドが入力されない状態が所定の時間以上継続した場合、ステップS151に進む。ステップS151において、エージェント制御部13は、エージェント172の表示を、所定の時間が経過する毎に、移動の状態、遊びの状態、または睡眠の状態に順次推移させる。

【0213】この待機中の処理の詳細について、図20のフローチャートを参照して説明する。なお、各ステップにおける処理は、エージェント制御部13が実行する。

【0214】ステップS161において、エージェント172の表示が、図24に示した待機中の状態から、例えば図33または図34に示した画像を用いて表現される移動の状態に推移する。

【0215】エージェント172の移動は、表示されているウィンドウと重畳しないようにデスクトップを横方向あるいは縦方向に行われる。なお、アクティブであ

るウィンドウ（いまの場合、メーラウィンドウ171）を抽出して、その周囲と横方向あるいは縦方向に行うようにしてもよい。エージェント172がデスクトップ上を横方向（例えば、右方向）に移動するときには、例えば、図33A乃至図33Dに示される画像が順次用いられることにより、瞬間移動したかのような動画表現が実現される。

【0216】具体的には、エージェント172の表示は、移動開始位置において、図33Aに示されるように、体の向きが移動する方向に向き、その後、向いている方向にジャンプすると、図33Bに示されるように、頭部から順に消滅して行く。そして、移動終了位置において、図33Cに示されるように、脚部から順に表示されて、最終的には、図33Dに示されるように全身が表示される。

【0217】エージェント172がデスクトップ上を上下方向に移動するときには、例えば図34A乃至図34Gに示される画像が順次用いられる。すなわち、移動開始位置において、エージェント172が、図34Aに示されるように、自身の尻尾（先端がコンセントプラグの形状をしている）を手で握り、図34Bに示されるように、尻尾の先端を頭上付近に差し込む。

【0218】その後、エージェント172の表示が、図34C、図34Dに順次示されるように、体の下部から徐々にロープに変身し、図34Eに示されるように、1本のロープになってその状態で移動終了位置まで移動する。移動終了位置においては、図34F、図34Gに順次示されるように、頭部から順に復元されて、最終的に全身が表示される。

【0219】このように、エージェント172の移動を、瞬間移動によって表現したり、1本のロープに変身させて表現したりすることにより、移動中を表現するために使われるリソース（演算量、メモリなど）の消費量を軽減させることが可能となる。

【0220】図20の説明に戻る。ステップS162において、イベント（電子メールの受信、電子メールの編集、あるいは関連情報の編集等を指示するコマンドの入力）が発生したか否かが判定される。イベントが発生していないと判定された場合、処理はステップS163に進む。

【0221】ステップS163において、エージェント172の表示が移動の状態に推移した後、所定の時間が経過したか否かが判定され、所定の時間が経過したと判定されるまで、ステップS162およびステップS163の処理が繰り返して実行される。ステップS163において、所定の時間が経過したと判定された場合、処理はステップS164に進む。

【0222】ステップS164において、エージェント172の表示が、移動の状態から、例えば図35に示される画像で表現される遊びの状態に推移する。図35A

は、エージェント172が蛇と戯れて遊んである状態を示しており、図35Bは、エージェント172が尻尾の先端を上方に差し込み、そこを支点としてぶら下がり揺れながら遊んである状態を示している。

【0223】ステップS165において、イベントが発生したか否かが判定される。イベントが発生していないと判定された場合、ステップS166に進む。ステップS166において、エージェント172の表示が遊びの状態に推移した後、所定の時間が経過したか否かが判定され、所定の時間が経過したと判定されるまで、ステップS165およびステップS166の処理が繰り返して実行される。ステップS166において、所定の時間が経過したと判定された場合、処理はステップS167に進む。

【0224】ステップS167において、エージェント172の表示が、遊びの状態から、例えば図36に示される画像で表現される睡眠の状態に推移する。ステップS168において、イベントが発生したか否かが判定され、イベントが発生するまで判定処理が繰り返して実行される。ステップS168において、イベントが発生したと判定された場合、実行されている待機中の処理は終了されて、処理は図19のステップS135に進み、それ以降の処理が実行される。

【0225】なお、ステップS162、またはステップS165において、イベントが発生したと判定された場合にも、実行されている待機中の処理は終了されて、処理は、図19のステップS135に進み、それ以降の処理が実行される。

【0226】また、図20のフローチャートには図示していないが、待機中の処理が実行されている最中において、メーラ2が終了されたと判定された場合にも、実行されている待機中の処理は終了されて、ステップS146に進む。同様に、ステップS133において、メーラ2が終了されたと判定された場合にも、処理はステップS146に進む。

【0227】ステップS146において、エージェント制御部113は、エージェント172の表示を、例えば図37Aおよび図37Bに示される画像で表現される、消失の状態に推移させる。図37Aは、エージェント172が手を振りながら背を向けて遠方に立ち去る状態を示しており、図37Bは、エージェント172の姿が徐々に小さくなり、やがて消失する状態を示している。

【0228】なお、エージェント172の消去とともに、吹き出し173、スクラップ帳ウィンドウ174、および推薦URL191等の表示も消去される。

【0229】以上のように、本発明によれば、電子メール等の文書から評価値の高い単語（重要語）を抽出し、関連情報を推薦する一連の処理に対応して、エージェント172が動作するので、エージェント172に対して信頼性や親しみが感じられるようになる。

【0230】ところで、上述したエージェント172の動作および吹き出し173の中の台詞の表示、並びに、表示された台詞に対応する音声信号の出力については、本発明のエージェントプログラム1だけでなく、他のアプリケーション、例えば、ゲームやワードプロセッサのヘルプ画面等に適用することが可能である。さらに、テレビジョン受像機、ビデオカメラ、またはカーナビゲーション等のディスプレイに表示されるキャラクタに適用することも勿論可能である。

【0231】また、同一のパーソナルコンピュータを複数のユーザが操作する場合、エージェント172の種類を複数用意して、ユーザ毎に表示されるエージェント172(図38)の種類を変えるようにしてもよい。また、エージェント172は、ユーザが好みのキャラクタを自由に作成し、編集できるようにしてもよい。

【0232】さらに、同一のユーザが複数のパーソナルコンピュータ上でエージェントプログラム1を利用する場合、異なるパーソナルコンピュータにおいても同じ種類のエージェント172が表示されるようにしてもよい。

【0233】なお、以上においては、エージェントプログラム1が実行されている場合、エージェント172は、常に登場しているものとして説明したが、例えば、推薦時にだけ表示させたりするように、その表示タイミングの設定を変更することができる。

【0234】具体的には、例えばエージェントプログラム1が実行されている状態において、マウスの右ボタンをクリックし、図38に示されるようなメニューボックス201を表示させて、その中から、「いろいろな設定をする」の項目を選択することにより、図39に示されるような設定画面を表示させる。

【0235】図39の例の設定画面には、複数のタブが配置されており、「エージェント」と示されたタブがアクティブとされているとき、ユーザが選択または入力可能な、エージェントの名前、表示、効果音、推薦間隔、推薦保存数、推薦するときの台詞、および推薦データ更新などの項目が表示される。

【0236】ユーザは、これらの項目に対して、それぞれ、所望の情報(エージェントの名前)を入力したり、あるいは、所定の項目を選択したりすることによって、自分好みにエージェント172および吹き出し173の表示状態、あるいは、推薦する関連情報の推薦間隔時間や保存数などを設定することができる。

【0237】次に、善積部11によるデータベースの更新のタイミングについて説明する。データベースは、上述したデータベース作成処理によって作成されるが、次のような第1乃至第3の状況になった場合、データベースが更新される。

【0238】すなわち、第1の状況として、データベースが作成または更新されてから所定の期間が経過してい

る場合、データベース内の関連情報が古くなってしまったので更新が行われる。

【0239】第2の状況として、データベースに蓄積されている関連情報のうちの所定の割合が提示済みとなった場合、データベース内の同じ関連情報が繰り返し提示されたり、提示する関連情報が不足したりしてしまうので更新が行われる。

【0240】第3の状況として、特徴抽出に用いた文書が電子メールである場合、電子メールの送受信が繰り返されていると、その文書の内容が変化するので更新が行われる。

【0241】なお、データベースの更新が必要である状況になった場合(例えば、イベント管理部31がタイマ31Aを監視し、所定の期間が経過したとき)、ユーザに対して更新を指示するように促すこともできるし、ユーザに対する更新指示の促しを実行することなく、自動的にデータベースを実行するように設定することも可能である。また、ユーザが指定する任意のタイミングで更新することも勿論可能である。

【0242】これら第1乃至第3の状況を考慮したデータベース更新処理について、図40のフローチャートを参照して説明する。このデータベース更新処理は、エージェントプログラム1が実行する処理のうちの1つであり、エージェントプログラム1の起動とともに開始され、エージェントプログラム1が終了されるまで繰り返し実行される。なお、この処理が開始される以前において、既に上述したデータベース作成処理が実行されており、データベースが存在するものとする。

【0243】ステップS181において、エージェントプログラム1の善積部11は、作成済みのデータベースの更新が必要であるかどうかを判定し、更新が必要であると判定されるまで待機する。この判断基準は、例えば図41に示すようなユーザインタフェースの画面を用いて、予めユーザが設定するものとする。図41の例では、4つの条件が示されており、ユーザによって左側の□印(チェックボックス)がチェックされた場合、対応する条件が有効となる。なお、1番目の条件では回数が設定可能とされており、3番目の条件では日数が設定可能とされている。

【0244】ステップS181において、更新が必要であると判定された場合、処理は、ステップS182に進む。ステップS182において、善積部11は、データベースを自動的に更新するように設定されているかどうかを判定し、自動的に更新するように設定されていないと判定した場合、ステップS183に進む。一方、ステップS182において、自動的に更新するように設定されていると判定された場合、ステップS183の処理はスキップされる。

【0245】ステップS183において、エージェントプログラム1の提示部12は、データベースの更新が必要

類である旨をユーザに通知するとともに、さらに、その通知に対応して、ユーザから更新の指示がなされたか否かを判定する。ユーザから更新の指示がなされたと判定された場合、処理はステップS184に進む。反対に、ユーザから更新の指示がなされないと判定された場合、処理はステップS181に戻り、以降の処理が繰り返して実行される。

【0246】ステップS184において、エージェントプログラム1の蓄積部11は、データベースを更新する。具体的には、文書取得部21乃至文書内容処理部23が、電子メールの電子メールボックスファイル（特定の拡張子`xxx`等が付与されていることが多い）を検出し、その更新日時を取得して、以前に取得した更新日時と比較し、異なる日付と異なるファイルサイズであれば、ファイルが更新されていると判断し、追加または変更された部分を抽出する。この場合、電子メールのグループ化、ヘッダの解析、形態素解析、特徴ベクトル算出等、一連のファイル内の分析が行われ、得られる重要語が関連情報検索部25に供給される。

【0247】ただし、メールグループ（話題）が変化せず（所定の話題に新たに追加された電子メールが少なく）、分析の結果、更新以前の重要語（検索用キーワード）と更新後の重要語が同じであれば、評価値等の計算値だけを変更し、関連情報検索部25による関連情報の検索を実行しないようにしてもよい。

【0248】あるいは、全ての電子メールグループが変化せずに一定期間が経過した場合、グループの特徴ベクトルのうち、前回、評価値が1番目と2番目の単語を検索語としていたもの、例えば評価値が3番目と4番目の単語を検索語に変更して検索し、検索結果を取得するようにしてもよい。

【0249】また、作り込み用単語組を用いた検索だけを行うようにして、データベースを更新するようにしてもよい。

【0250】なお、関連情報をインターネット上の検索エンジンを用いて検索する際、インターネットに接続している状態であるか否かを検出するようにし、インターネットに接続していない状態である場合、関連情報の検索を行わないようにし、以降においてインターネットに接続した状態となったときに関連情報を検索するか否かをユーザに問うようにしてもよい。

【0251】「同一関連情報を何度も推薦（提示）しないようにするために、あるメールグループの関連情報を、所定の回数を以上推薦したら更新が必要と判断する」との条件に関連して、取得した電子メールと類似性の高いメールグループ（話題）を選択する際に、同じメールグループから何度も推薦を行わないように、次のような処理を行う。

【0252】メールグループ自体に推薦の優先度の順位を付与し（例えば、メールグループ内での特徴語の評

値の最大値をそのメールグループの優先度の値とし、優先度の値を降順に並べたものを優先度の順位として付与する）、一度推薦を行ったメールグループを優先順位の最後尾に並び替えるようにする。このようにすることによって、類似度の範囲内にあるメールグループでも、同じメールグループから推薦する頻度が減少する。また、優先順位の変更だけなので、関連情報を大量に検索して準備しておけば、なるべく同一メールグループからの推薦が減り、かつ、情報自体も不足することなく用いることができる。

【0253】これに関連して、特徴抽出に用いる話題内の文書量に応じて、類似する話題を抽出する際の範囲を変化させることができるようにする。具体的には、特徴抽出する話題の文書量またはデータサイズに応じて何段階かの類似度の範囲を設定する。例えば、ある話題に含まれる文書量が10ファイル以内である場合は類似度を0.01以上、11ファイル以上50ファイル未満の場合は類似度を0.03以上、5150ファイル以上である場合は類似度を0.05以上とする。または、ある話題の文書の容量が500キロバイト未満である場合は類似度を0.01以上、500キロバイト以上である場合は類似度を0.02以上とする。

【0254】そして、予め設定された類似度の範囲のうち、優先度の高い話題から検索された関連情報を提示するようにする。このようにすると、文書量の減少により、データベースの内容が更新されると、類似度の範囲が変化し、類似度の範囲が狭まって関連情報が不足したり、反対に、類似度の範囲が広くなってユーザにとってあまり関連性が明確でない関連情報が提示されたりするような事態の発生を抑制することが可能となる。

【0255】以上説明したように、データベース更新処理においては、追加された文書や変更された文書だけを処理の対象とするので、データベース作成処理を繰り返して実行する場合に比べて、処理時間が短縮される。

【0256】本発明のエージェントプログラム1は、上述したようにメール2によって送受信される電子メールやワードプロセッサプログラム3で編集される文書の他、例えば、チャット、電子ニュース、電子掲示板等の文書や音声信号をテキスト化した文書など、属性情報としてタイムスタンプが付与されている文書に対応して動作するようにさせることができる。

【0257】上述した一連の処理を実行するエージェントプログラム1は、パーソナルコンピュータに予め組み込まれるか、あるいは、記録媒体からインストールされる。

【0258】上述した一連の処理は、ハードウェアに実行させることもできるが、通常、ソフトウェアにより実行させる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するエージェントプログラム1が、専用のハードウェアに組み込まれている

コンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、記録媒体からインストールされる

【0259】コンピュータにインストールされ、コンピュータによって実行可能な状態とされるプログラムを記録する記録媒体は、図2に示されるように、プログラムが記録されている磁気ディスク52（フレキシブルディスクを含む）、光ディスク53（CD-ROM（Compact Disc-Read Only Memory）、DVD（Digital Versatile Disk）を含む）、光磁気ディスク54（MO（Mini-Disk）を含む）、もしくは半導体メモリ55などよりなるパッケージメディア、または、プログラムが一時的もしくはが永続的に記録されるROM42や記憶部49を構成するハードディスクなどにより構成される。記録媒体に対するプログラムの記録は、必要に応じて、ルータ、モデムなどのインタフェースを介して、公衆回線網、ローカルエリアネットワーク、インターネット、ディジタル衛星放送といった、有線または無線の通信媒体を利用して行われる。

【0260】なお、本明細書において、記録媒体に記録されるプログラムを記述するステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。

【0261】

【発明の効果】以上のように、本発明によれば、速やかにユーザの興味に対応する単語を抽出し、電子メールの送受信が行われない状況においても、ユーザに適切な情報を提示することが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施の形態であるエージェントプログラムの機能ブロックの構成例を示す図である。

【図2】エージェントプログラムをインストールして実行させるパーソナルコンピュータの構成例を示すブロック図である。

【図3】エージェントプログラムによるデータベース作成処理を説明するフローチャートである。

【図4】図3のステップS5の処理を説明するための図である。

【図5】図4のステップS22における、日時条件およびアドレス属性条件を設定処理説明するフローチャートである。

【図6】話題ファイルの一例を示す図である。

【図7】単語ベクトルを構成する複数の単語に含まれる要素を示す図である。

【図8】図3のステップS3における第1話題選択処理を説明するフローチャートである。

【図9】図3のステップS4における形態素解析処理を説明するフローチャートである。

【図10】話題単語テーブルの構成例を示す図である。

【図11】単語インデックステーブルの構成例を示す図である。

【図12】話題評価値テーブルの構成例を示す図である。

【図13】図3のステップS5における不要語削除処理を説明するフローチャートである。

【図14】図3のステップS9における第2話題選択処理を説明するフローチャートである。

【図15】図3のステップS11における推薦話題確定処理を説明するフローチャートである。

【図16】ステップS12におけるWeb検索処理を説明するフローチャートである。

【図17】エージェントプログラムの関連情報提示処理を説明するフローチャートである。

【図18】図5のステップS15の処理を説明するための図である。

【図19】エージェントの動作等を説明するフローチャートである。

【図20】図7のステップS51の待機中の処理の詳細を説明するフローチャートである。

【図21】デスクトップ上に表示されたエージェントの表示例を示す図である。

【図22】エージェントが登場するときの表示例を示す図である。

【図23】エージェントの台詞である吹き出しの表示例を示す図である。

【図24】エージェントが待機中であるときの表示例を示す図である。

【図25】エージェントが作業中であるときの表示例を示す図である。

【図26】デスクトップ上に表示された入力ウィンドウの表示例を示す図である。

【図27】入力ウィンドウの表示例を示す図である。

【図28】デスクトップ上に表示された推薦URLの表示例を示す図である。

【図29】エージェントが指示中であるときの表示例を示す図である。

【図30】デスクトップ上に表示されたスクラップ帳ウィンドウの表示例を示す図である。

【図31】エージェントが遊びの状態であるときの表示例を示す図である。

【図32】エージェントが悲しみの状態であるときの表示例を示す図である。

【図33】エージェントが横方向に移動するときの表示例を示す図である。

【図34】エージェントが縦方向に移動するときの表示例を示す図である。

【図35】エージェントが遊びの状態であるときの表示例を示す図である。

【図36】エージェントが睡眠の状態であるときの表示

例を示す図である。

【図37】エージェントが立ち去るときの表示例を示す図である。

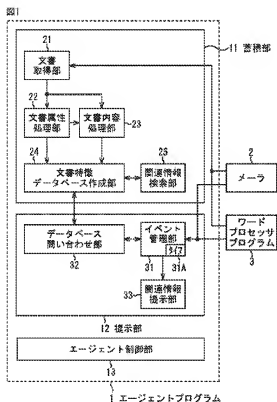
【図38】メニューボックスの表示例を示す図である。

【図39】設定画面の表示例を示す図である。

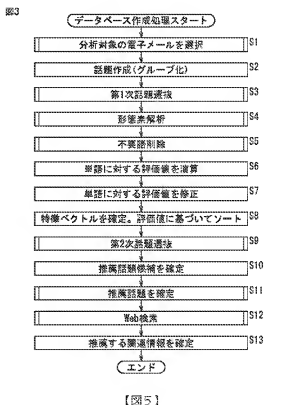
【図40】エージェントプログラムのデータベース更新処理を説明するフローチャートである。

【図41】データベースを更新させる条件を入力するユーザインタフェースの表示例を示す図である。

【図1】



【図3】



【図7】

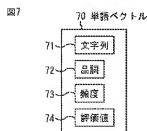
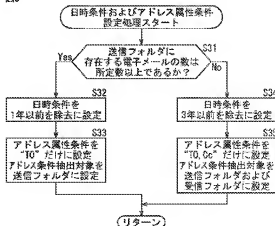
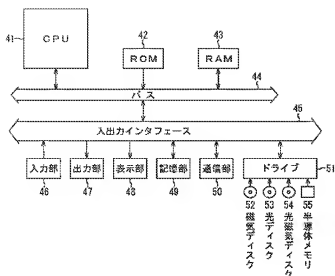


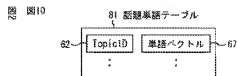
図5



【図2】



【図10】

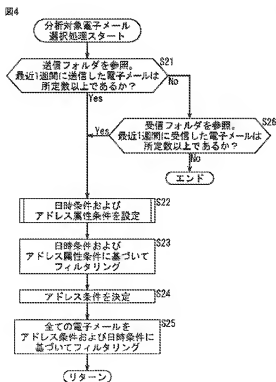


【図24】

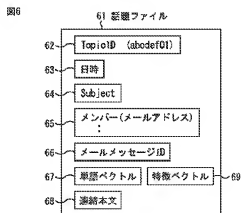
図24



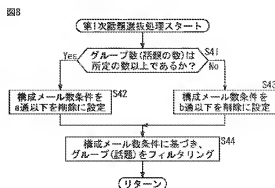
【図4】



【図6】

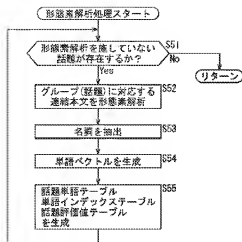


【図8】



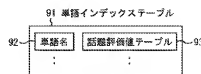
【図9】

図9



【図11】

図11



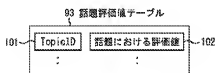
【図25】

図25



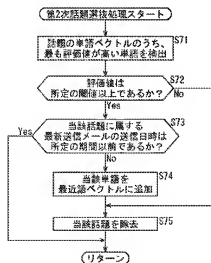
【図12】

図12



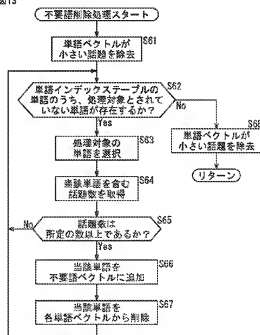
【図14】

図14



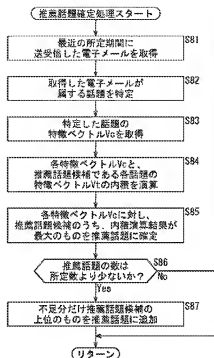
【図13】

図13



【図15】

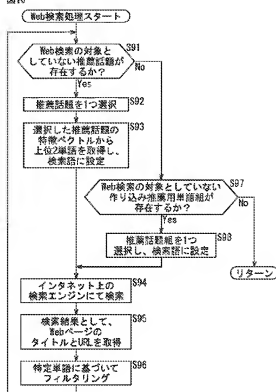
図15



【図18】

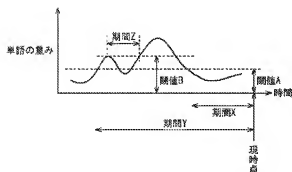
【図16】

図16



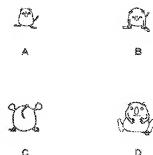
【図22】

図 22



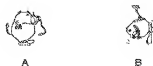
【図36】

図36



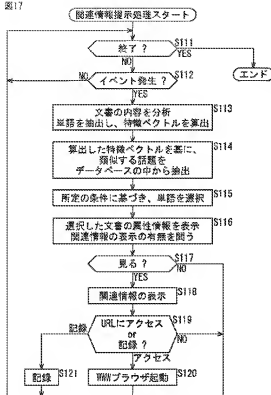
【図35】

図35

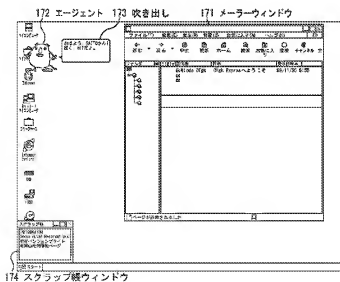


【図17】

図17

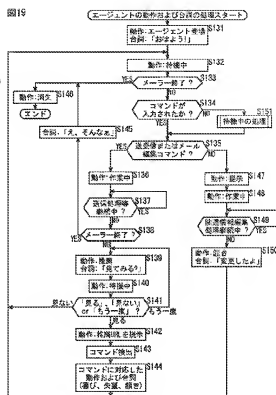


【図21】



【図19】

図19



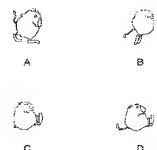
【図32】

図32



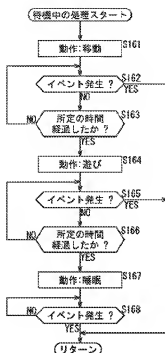
【図33】

図33



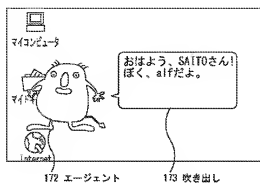
【図20】

図20



【図23】

図23

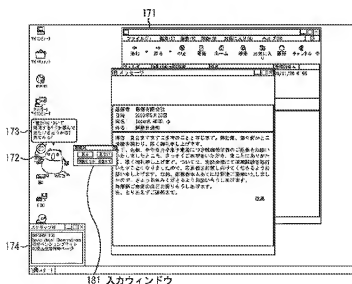


【図31】

図31



【図26】



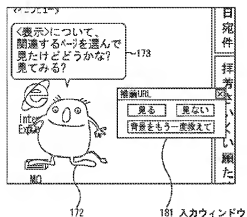
【図34】

図34



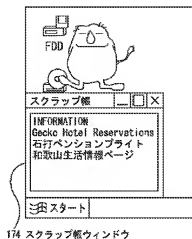
【図27】

図27



【図30】

図30



【図28】

【図37】

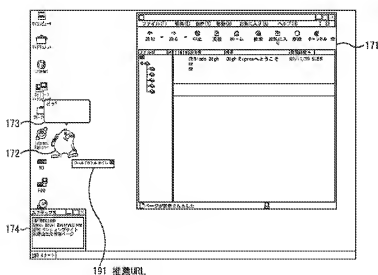
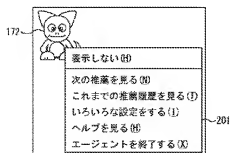


図37

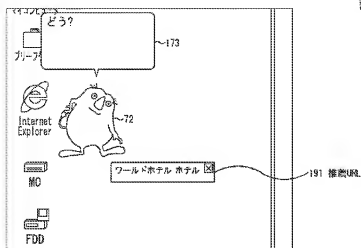


【図38】

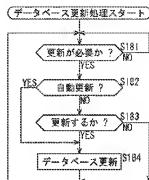
図38



【図29】

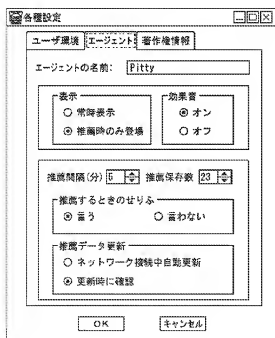


【図40】



【図39】

図39



【図41】

図41

- ☐URLが 回行われたメールグループは更新必要と判断する
☐添装は使ったメールグループは更新必要と判断する
☐最終更新から 日経過したメールグループは更新必要と判断する
☐新着メールの分析により、検索語が変更されたメールグループは更新必要と判断する

フロントページの続き

(51)Int.Cl.?

G 0 6 F 13/00

識別記号

6 2 0

FI

G 0 6 F 13/00

フィード (参考)

6 2 0